The Dissertation Committee for Po-Ting Chen Certifies that this is the approved version of the following dissertation:

Typicality in Chinese Sentence Processing: Evidence from Offline Judgment and Online Self-paced Reading

Committee:

Richard P Meier, Supervisor

Nicole Y Y Wicha, Co-Supervisor

Colin J Bannard

Catharine H Echols

Zenzi M Griffin

Shukhan Ng

Stephen M Wechsler

Typicality in Chinese Sentence Processing: Evidence from Offline Judgment and Online Self-Paced Reading

by

Po-Ting Chen, B.A.; B.B.A; M.A.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August 2014

Dedication

This dissertation is dedicated to my parents, Jack C.K. Chen and Grace H. Huang, for your unfaltering and unconditional love and support.

Acknowledgements

I embarked on this PhD journey nine years ago. As I look back at this long but fulfilling journey, I would not have accomplished my goal without love, support and guidance from so many wonderful people.

I owe my deepest gratitude to my dissertation committee, especially my supervisors Richard Meier and Nicole Wicha, both who have been the most patient and kind supervisors one could ask for. Their guidance, expertise, generosity of their time, and encouragement were crucial in helping me overcome the challenges that I encountered along the way. I am thankfully for all the time that I spent at the Wicha Lab in UT-San Antonio, where Nicole Wicha and Shukhan Ng were always generous with time, feedback, and made the countless trips to San Antonio immensely valuable. I'm also thankful for the contribution from my other committee members. Steve Wechsler, who was also my reader for my master's thesis, provided expertise and suggestions that helped strengthen my work. Colin Bannard, provided invaluable statistical advice and helped me understand how to conduct quantitative analysis, prompting me to develop a wide range of statistical analysis skills. I'm also grateful for my committee members from the Psychology department, Cathy Echols and Zenzi Griffin, who provided insightful and crucial feedback to my data analysis.

I'm grateful to the Linguistics and Statistics departments at UT for the financial support I've received. The TAships and AIship not only provided crucial financial support but they also led me to discover my passion for teaching.

I would not have been able to collect my data without the generous support of my mentor at NCCU in Taiwan, Cynthia Tsui, who has twice graciously gave me access to her students. With her help and support, I was able to recruit sufficient participants and

conduct my experiments smoothly. I'm also deeply grateful for all the friends in Taiwan and around the world that helped me fill out the questionnaires and recruit participants. I'm grateful to Mike Dobberfuhl for helping me set up the online experiments for me. His programming expertise was crucial in ensuring that my experiments were setup correctly.

I'm very grateful to be surrounded by a wonderful group of friends I have met here in Austin. My closest friends, especially Wendy Kan and Qiuana Lopez, who I shared lots of ups and downs with, were always full of great advise, have given me invaluable encouragement and always pushed me beyond what I believe I could do. I'm also thankful for my friends, Gaby Garcia, Justin McIntosh, Shawn Chang, Niamh Kelly, Rachael Gilbert, Lauren Terzenbach, Ya-lan Chang, Pei-yin Chien, Hsiangyun Chen, Lydia Baldwin, and many more, who made my life in Austin an absolutely joyful one,

Lastly, I'm deeply grateful for having a supportive family. My mum, Grace Huang, my dad Jack Chen, my brother Charlie Chen, and sister-in-law Wen-Ting Chen are the kindest and most positive people. I'm indebted to the generous support I've had for them and for the encouragement they've given me. And finally, I'm deeply grateful for Brian Byrne, who joined me at the last but the most crucial part of the journey and provided me with the utmost love and support that I could have asked for.

Thank you all, I couldn't have done it without you.

Typicality in Chinese Sentence Processing: Evidence from Offline Judgment and Online Self-paced Reading

Po-Ting Chen, Ph.D.

The University of Texas at Austin, 2014

Supervisors: Richard P Meier, Nicole Y Y Wicha

This study examines how Chinese speakers understand sentences describing events that have varying degrees of typicality. How the interpretation of typicality is obtained from linguistic input is not fully understood. In this study, I investigate the association of pairs of content words in order to determine their contribution to judgments of event typicality. The associations between words could influence the interpretation of event typicality. Two words that are not associated semantically, for example *baby* and *wine,* may be seen as an atypical combination. However, when these words are placed in a sentence context, the resulting sentences can be a typical scenario, such as *the baby spilled the wine.*

Four offline judgment studies were conducted to obtain quantitative measurements of the association of word pairs and of judgments of event typicality in sentences. These studies demonstrated that noun pairs showed larger differences in their association ratings than those of noun-verb pairs. When the sentences containing the word pairs were judged, the association of the noun pair strongly influenced the sentence's event typicality ratings, regardless of word order or of the typicality of the

verb. Two online, word-by-word self-paced reading studies were conducted to examine whether judgments of word associations and event typicality are used in real-time sentence processing. The results showed that there was a slowdown in reading times at the critical regions when the noun pairs were atypical. The typicality of the verb did not result in a difference in reading times, regardless of the word order of the sentences, although offline judgment scores of event typicality were predictive of online reading times.

The findings of these studies suggest that: (1) event typicality is more than the semantic association between words. Noun-noun and noun-verb associations contribute to event typicality but the association of two nouns has a more significant contribution and is not affected by an intervening word, (2) the typicality of verbs contributed to real-time sentence processing, insofar as the verbs contributed to the judged typicality of the events expressed by SVO and SOV clauses, and (3) in real-time sentence processing, regardless of the sentence's word order, the association of nouns has greater impact on event typicality processing. This is not likely to be due simply to a priming effect between nouns, but rather also reflects the processing of the sentence's event typicality.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

1.1 INTRODUCTION

Based on our intuition and real world knowledge, when we hear the words *the chef washed* we are likely to think of continuation words such as *knife, cucumber,* or any item that is typicality used by or associated with a chef. A possible continuation of the phrase could, however, be *the chef washed the military uniform*. While this is a plausible event it is not a typical one, because we do not typically associate *chef* with *military uniform*.

The above example exemplifies our understanding of a typical event and the relationship between two words. Two words can be related in various ways: in form, function, or in meaning. The words *chef* and *knife* are related in meaning and have a strong association. On the other hand, the words *chef* and *military uniform* are weakly associated. Regardless of the strength of association between two words, they can be placed in a sentence context to create events that vary in their plausibility. There is, for example, a strong association between the words *chef* and *knife*. When these words are embedded in a sentence, *the chef washed the knife* is a plausible event, while *the knife washed the chef* is an implausible event. We know *the knife washed the chef* is implausible because there is a selectional restriction violation in this sentence; *knife* is an inanimate entity and cannot perform the action of washing.

Plausible and implausible events can usually be distinguished by the presence of selectional restriction violations, whereby sentences expressing plausible events do not have such violations. Within any given set of plausible sentences, there is a continuum of events that have different likelihoods of occurrence, where some events are highly typical and some events are less typical. The relationship between event plausibility and event typicality is illustrated in Figure 1.1 below.

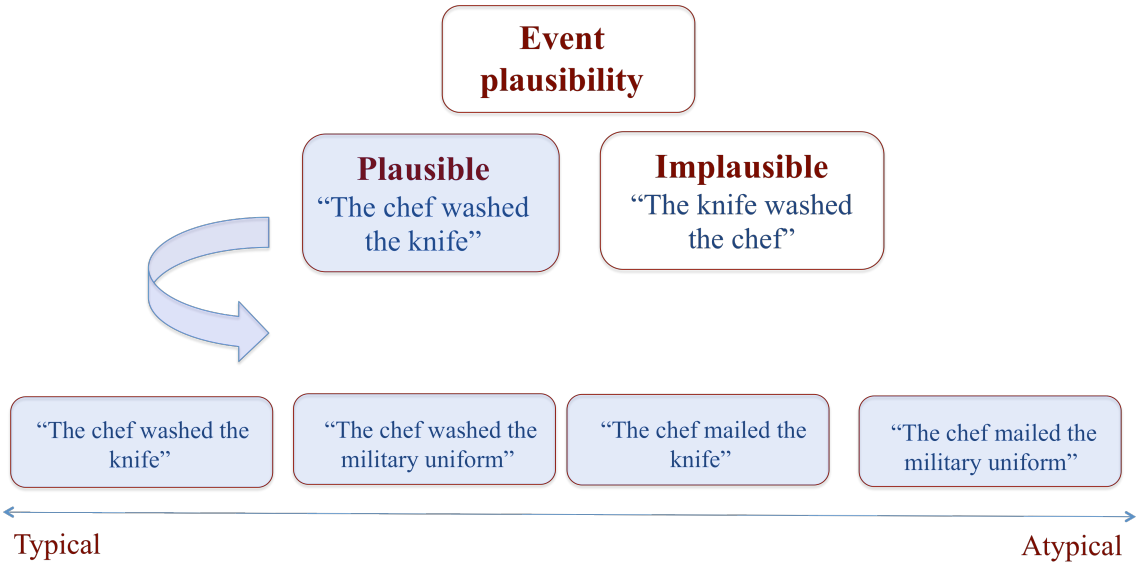

Figure 1.1: A schematic representation of event plausibility and event typicality.

The examples shown in Figure 1.1 shows two strongly associated words can create different events that vary in typicality. *The chef washed the knife* may be perceived as more typical than *the chef mailed the knife*. Alternatively, two words that may not appear

to be associated, such as *chef* and *military uniform,* can be placed in a sentence context to create a more typical scenario, such as *the chef washed the military uniform*. By placing two words in a sentence context, we can determine whether the association between the two words can contribute to event typicality.

The examples mentioned thus far show the relationship between the two concepts: the semantic association between words and event typicality. When judging the typicality of an event expressed by the sentence, how do the association between words contribute to our judgments? The goal of this dissertation is to understand how the semantic association of words contributes to event typicality and whether event typicality affects real-time sentence processing. In particular, this dissertation looks at different word combinations and sentences with different word orders in Mandarin Chinese in offline and online studies to investigate how individual linguistic input contributes to event typicality judgments. The flexibility of word order in Mandarin Chinese and its simple morphology, allows a better assessment of the effects of event typicality than would be true in many other languages. These properties of Chinese will be discussed further in the next chapter.

1.2 REAL WORLD KNOWLEDGE AND LINGUISTIC KNOWLEDGE

One of the ways that we can distinguish events that vary in typicality is by using our knowledge of real world events. The integration of real-world knowledge in sentence processing has received the attention of psycholinguists in the last few decades. This

knowledge is considered one of the key factors in human language processing (e.g., Bicknell et al., 2010; Cook & Myers, 2004; Filik, 2008; Hagoort et al., 2004; Matsuki et al., 2011, McRae et al., 1998; Metusalem et al., 2012; Staub et al., 2007; Warren et al., 2008; Van Berkum et al., 2005). Real world knowledge has been studied to see how people use it to form an association between two words (Hare et al., 2009), to compare the truth-value of a sentence with real world events (Hagoort et al., 2004), to identify plausible and implausible events (Kutas & Hillyard, 1980, 1984), and to distinguish events based on their typicality, or their likelihood of occurrence in the real world (Bicknell et al., 2010; Matsuki et al., 2011, Zarcone & Pado, 2011).

Real world event knowledge is knowledge that is shared among people, where people have a generalized or prototypical representation of common events (McRae & Matsuki, 2009). Using this knowledge, we can interpret the likelihood of the occurrence of an event, such that events that are more likely to occur are considered typical and those that are less likely are considered atypical. When we hear the sentence *the kids ate Brussels sprouts at the birthday party,* for example, we are likely to infer that this is not a typical event, especially if we compare it to the more typical scenario *the kids ate cake at the birthday party*. Although both scenarios are plausible, we make typicality judgments based on what we know about children's eating preferences and birthday parties.

A speaker's social and cultural background will likely affect the judgment of an event's typicality. For example, when we think of the event *going to a children's birthday party,* we think of the items and activities that are related to this event, such as eating a birthday cake, playing games, and opening presents. While the specific types of

cake and games may differ, there is a general understanding of what characterizes typical events at such a party among people of similar social and cultural backgrounds. *The kids hit a piñata*, for example, would be a plausible and typical activity at a child's birthday party in the US, whereas it would be plausible but not typical in Asia. *The kids ate birthday cake* would be a typical event at a child's birthday party in both the US and Asia.

Interpreting the meaning of a linguistic input involves using both our linguistic knowledge and our real world knowledge. Constructing meaning is essential, but how do we accomplish this from the linguistic input? Some have argued that meaning is encoded in the linguistic units that are stored in the lexicon, for example: the verb *eat* encodes the requirement for the agent as [+animate] and the patient noun to be [+edible] (Chomsky, 1965). On the other hand, we also bring in knowledge of real world events, and this knowledge is not explicit in the lexicon (Hagoort et al., 2004). It is not always obvious whether there really is a clear line between knowledge that is embedded in the lexicon and knowledge that is not (Jackendoff, 2002).

Jackendoff (2002) presents an argument whereby he suggests that the boundaries between lexical and world knowledge are blurry, and the existence of such boundaries is questionable. In his study he presented examples such as *John finished the book* versus *the goat finished the book,* where we know the verb *finished* can imply the completion of different actions – reading, eating, and destroying, depending on the identity of the doer and the affected object. We also know that the agent *John,* as a human being, is less likely to eat a book than is a goat. In such cases, there appears to be no distinct

boundaries between semantic, contextual, and real world knowledge. Whether or not real world knowledge is stored in the lexicon is not the main focus of this dissertation; instead, real world knowledge is viewed as the context in which the interpretation of linguistic input takes place.

1.3 EVENT PLAUSIBILITY AND EVENT TYPICALITY

Real world knowledge is informative, it helps us to evaluate the relationship between two words, determine a sentence's meaning, and detect meaning violations. Real world knowledge can inform us about the association between words. Priming studies have shown that people use their real world knowledge to associate words that might otherwise be unrelated, for example the noun pair, *chef* and *knife* (Hare et al., 2009), and the subject-verb pair, *waiter* and *serving* (McRae et al., 2005). We know chefs typically use knives and waiters serve; therefore we form a strong association between *chef* and *knife* and between *waiter* and *serving*. As well as being able to help us judge the relationship between two words, real world knowledge allows us to assess the plausibility and typicality of events.

Real world knowledge helps us to evaluate a sentence's meaning and detect meaning violations. A meaning violation can be semantic or pragmatic. A semantic violation arises when the meaning of the sentence is implausible or anomalous,[1] with

[1] Some studies make a further distinction between implausibility and semantic anomaly, e.g., Rayner et al., 2004. However, Murray (2006) disagrees since this distinction is not always clear. Implausible and

such sentences incurring a higher processing cost (Ferreira, 2003; Hagoort, 2004). These sentences usually have a selectional restriction violation, for example, *the cheese ate the mouse* is implausible, because the verbal argument *cheese* does not meet the verb's animate requirement for its agent role. Conceptually, we know this event is unlikely to happen in the real world. A pragmatic violation, on the other hand, arises when the expressed meaning is not consistent with a real world event, or when the sentence expresses an event that is plausible but not typical.[2] Sentences with pragmatic violations express events that are plausible, but that may differ in their likelihood of occurrence. They also do not pose any selectional restriction violations, and thus the distinction between a typical and an atypical event may rely heavily on real world knowledge.

Hagoort et al. (2004) showed that a sentence with pragmatic violation, such as *Dutch trains are white,* elicited an N400 effect, a detection of meaning anomaly, because Dutch trains are in fact yellow. For Dutch speakers, this is a pragmatic violation; for people who are not familiar with Dutch trains, the violation in this sentence may not be apparent. Plausible events can differ in their typicality. Bicknell et al. (2010) showed that plausible sentences like *the mechanic checked the brakes* and *the journalist checked the brakes* were processed differently; the first sentence was more typical and showed neither an N400 effect nor a slowdown in reading times, whereas the second showed both kinds

---

semantically anomalous sentences are not the focus of this dissertation and not distinguished here. This dissertation focuses on typical and atypical events that are plausible, and an atypical event is one that is pragmatically anomalous because it is less likely to happen.

[2] Murray (2006) calls these events a 'more subtle' violation of plausibility, such as *the duckling stretched the smoking cardigan* where the event is not implausible but simply less likely to happen.

of effects. The processing difference was attributed to the difference in the likelihood of occurrence of the events, since neither sentence had a selectional restriction violation. Although the arguments were good fits for the verb, *mechanic* was a better fit than *journalist*. The authors argued that the evaluation was based on knowing the verb's structure as well as the fit of the verbal arguments with the verb. Decisions as to the thematic fit of the verbal arguments are determined by real world knowledge.

The typicality of an event is determined by how likely an action is performed by a person (Bicknell et al., 2010), and whether an instrument is typically used to perform a given action (Matsuki et al., 2011). These studies are informative in showing that individual words do not determine whether the sentences are typical or atypical; rather it is the combination and relatedness with other words that create a sentence meaning that can be evaluated for event typicality (Bicknell et al., 2010; Matsuki et al., 2011). However, the picture of how people form judgments of typicality, and more specifically how the association between two words may affect event typicality judgment, is still not fully clear. People formulate associations between two words based on real world knowledge. The interpretation of an event in a sentence is derived from understanding the relationship among words; therefore, the association between two words may influence the assessment of event typicality.

1.3.1 Semantic association and event typicality

To understand how people evaluate the typicality of the event communicated by a sentence, we need to consider the relationship between two words, because the

relationship between two words can affect a sentence's meaning. In this dissertation, the semantic association or relationship between two words will be considered.

Two words can be related in various ways, whether in form or in meaning. Studies have found that in lexical decision tasks, participants were faster to make a decision on whether two strings were real words when they were taxonomically associated, such as *car-truck*, and semantically related[3] such as *bread-butter* than words that were not related or are not associated, such as *bread-stem* (e.g., Fischler, 1977; Lupker, 1984).

The interpretation of the association between words could also be changed by discourse context (Filik and Leuthold, 2008; Nieuwland & Van Berkum, 2006). The words *peanut* and *fall in love* may not appear to have a strong association; in fact, the juxtaposition of these words forms an anomalous meaning. However, when placed in an appropriate discourse context where the peanut was anthropomorphized, an interpretation of *peanuts fall in love* can be formed and accepted (Nieuwland & Van Berkum). Similarly, the strong association of two words such as *peanut* and *salted* was weakened when the preceding context of an anthropomorphized peanut did not favor a *salted peanut* interpretation. On the other hand, two words may have a strong association that cannot be suppressed by sentence context. Fischler et al. (1983) showed that the N400 amplitude was similar when participants heard the sentences *a robin is a bird* and *a robin is not a bird* and this amplitude was much smaller than the sentences *a robin is a vehicle*

[3] Fishcler (1977) and Lupker (1984) refers to the pair *dog-cat* as an associated pair and *bread-butter* as a semantically related pair. Markman and Hutchinson (1984) refers to *dog-cat* as a taxonomically related pair and *bread-butter* pair as a thematically related pair.

and *a robin is not a vehicle.* The authors attributed this N400 effect to the mismatch of the nouns *robin* and *vehicle*, rather than the truth-value of the sentences. This finding provides some evidence that a strong association between two words cannot be suppressed by sentence context, at least not initially.

In this dissertation, I examined whether an association between words can contribute to the judgment of event typicality, since the role of word association in event typicality assessment is not fully understood. The study of Nieuwland & Van Berkum showed that the plausibility of an event is not simply determined by the association between two words; the discourse context can override the strong association between words. The goal of understanding event typicality judgment will be accomplished by examining word associations in Chinese. Chinese word pairs and sentences were used because of the availability of various word orders, including SOV. The relationship between two nouns was examined first. The noun pairs were then placed in a sentence context, where the association of the verb with the noun pair was manipulated to create events that varied in typicality. In a sentence context, if the association between two words contributes to event typicality, this shows that lexical association overrides event typicality. On the other hand, if the association between words does not contribute to event typicality, this shows that event typicality overrides lexical association.

1.4 GOALS AND RESEARCH QUESTIONS

The goal of this dissertation is to understand whether event typicality affects sentence processing. This will be accomplished by examining how the association between words contributes to event typicality and how event typicality unfolds in real-time sentence processing. In addition, the following questions will be answered:

1. Is event typicality more than the semantic association between words?
2. Does event typicality affect real-time reading?

The first question will address the relationship between word association and event typicality, to determine whether judgments of word associations contribute to event typicality. Judgments of word pair associations were obtained as well as judgments of sentences that depict event of varied typicality. Strongly associated and weakly associated noun-noun and noun-verb pairs were created and then placed in sentences with different word orders. This allows the contribution of word to event typicality to be examined.

The second research question addresses whether judgments of word associations and event typicality are used in real-time sentence comprehension. By measuring reading time at critical points of a sentence, we can compare sentences with varying degrees of event typicality to examine the impact of word association. Word order was manipulated to determine if the association of word pairs is sensitive to word order.

Two types of studies were conducted to address the research questions: four offline judgment studies and two online word-by-word self-paced reading studies. The

judgment studies were conducted to elicit Chinese speakers' intuitions about the association of the word pairs – specifically their intuitions about whether or not two words were strongly or weakly associated – as well as their judgments of the typicality of the events expressed by a sentence. The purpose of the judgment studies was to obtain measurements of word association and sentence event typicality. Word pairs and sentences may differ in their association and there are different ways this association can be determined. In this dissertation, I will measure this association based on judgments of their likelihood of occurrence, where Chinese speakers will be asked to judge word pairs and event typicality on a seven-point scale. In both sets of studies, the association of words within noun pairs was manipulated. In a subject-object pair, the second noun was manipulated for its association with the first noun, resulting in a strongly associated or weakly associated noun pair. This noun pair was then combined with a verb, where the association of the verb with the noun pair was manipulated, resulting in sentences expressing events of varying degrees of typicality. These manipulations created typical and atypical word pairs and events, thus allowing the typicality of the events to be assessed as well as allowing us to probe the contribution of word association to event typicality.

The offline subject-object pairs and subject-object-verb sentences suggested that, the association between two nouns had more significant contribution to event typicality. To confirm this observation, follow-up studies were conducted to further determine the contribution of verbs in event typicality judgment, specifically whether this effect of noun typicality could be affected by having intervening words between two nouns.

Results of the judgment studeies provided the basis for setting up subsequent self-paced reading studies, whih were used to determine whether the offline judgments of word pairs and event typicality modulate sentence comprehension in real-time. In the first self-paced reading study participants read SOV sentences. A strong effect of subject-object association was present and a follow-up reading study with SVO sentences showed similar results. Regression analysis showed that reading time was predicted by offline SOV and SVO judgment scores, indicating an effect of event typicality in real-time sentence processing.

1.5 OUTLINE OF THE DISSERTATION

The outline for the chapters in this dissertation is as follows: Chapter two provides background on event plausibility and event typicality in sentence processing. Chapter three reports the results of four judgment studies in which Chinese speakers evaluated the association of word pairs and sentences expressing events that vary in typicality. Chapter four reports two online self-paced reading studies that examined the contribution of word association and event typicality in real-time sentence processing. Finally, chapter five summarizes the results from chapters three and four and discusses the implications of the findings in the context of sentence processing.

Chapter 2 Background and literature review

2.1 INCREMENTAL SENTENCE PROCESSING

Sentence processing occurs rapidly and incrementally (Bornkessel & Schlesewsky, 2006, Marslen-Wilson, 1975, Tanenhaus et al., 1995); the process of constructing a meaningful interpretation occurs on a word-by-word basis as the sentence unfolds. In order to create a coherent meaning for a sentence, both syntactic and semantic knowledge are accessed and integrated simultaneously (e.g., Kamide, Altmann, & Haywood, 2003; King & Just, 1991; Tanenhaus, Carlson, & Trueswell, 1989; Traxler, Morris & Seely, 2002). In addition, people draw on their pragmatic knowledge of real world events in real-time sentence processing (e.g., Altmann & Kamide, 2007; Bicknell et al., 2010; Hare et al., 2009; Hagoort et al., 2004; Matsuki et al., 2011; Metusalem et al., 2012; McRae et al., 2005, McRae, Spivey-Knowlton, & Tanenhaus, 1998). The integration of real world knowledge in real-time sentence processing has been used as evidence to question processing models suggesting that syntactic information is privileged and is accessed prior to non-syntactic information (Ferreira & Clifton, 1986; Frazier, 1995; Rayner et al., 1983; Van Gompel et al., 2005).

The integration of real world knowledge in language processing has been studied using different methods. Event-related-potential (ERP) studies (e.g., Federmeier and Kutas, 1999; Kutas & Hillyard, 1980; Van Berkum et al., 2005) and eye-tracking studies (Altmann and Kamide, 2007; Camblin et al. 2007; Matsuki et al., 2011) have provided

evidence demonstrating that expectations based on real world knowledge are computed immediately without delay during sentence comprehension. Real world knowledge also allows us to distinguish plausible and implausible events (Hagoort et al., 2004), and events that differ in their likelihood of occurrence (Bicknell et al., 2010). In this dissertation, the focus is on judge the typicality of the event expressed by the sentence, which may be based on real world knowledge, linguistic knowledge or both.

2.2 REAL WORLD KNOWLEDGE

Real world knowledge is crucial in sentence processing because it provides a context by which we can interpret the meaning of sentences. This knowledge is the accumulation of our everyday experience, forming the basis of our knowledge of generalized or prototypical event types (McRae & Matsuki, 2009). Having this contextual information allows us to judge the plausibility and typicality of the event expressed in a sentence.

One of the ways that event knowledge has been studied is by examining the interpretation of implausible sentences (e.g., Hagoort et al., 2004, Kutas & Hillyard, 1980, 1984). An anomalous sentence can be one where the truth-value denoted by the sentence contradicts real world events, or where the meaning is semantically incoherent. Hagoort et al. showed that when Dutch participants heard the sentence *the Dutch trains are white,* a N400 effect, a detection of meaning anomaly, was elicited because Dutch trains are yellow. A N400 effect is an event-related-potential where a negative deflection

occurs around 400 milliseconds after the onset of the stimulus word (Coulson et al., 1998). While the sentence is grammatical, the meaning is not consistent with the participants' understanding of the real world. A contrasting sentence that expresses a semantically anomalous idea, *the Dutch trains are sour* also elicited a N400 effect.

Linguistically, making a distinction between plausible and implausible events can largely be determined by selectional restriction violations. If a sentence has a selectional restriction violation, it is highly likely to result in a semantically anomalous meaning. Sentences expressing plausible scenarios on the other hand, have no such violations. Ferreira (2003) showed in sentences that express anomalous ideas, where the expected agent and patient roles were reversed such as *the cheese ate the mouse*, participants were less accurate at identifying the correct thematic role (agent or patient) compared to its plausible counterpart *the mouse ate the cheese.* While both of these sentences are grammatical, the former is a highly unlikely event since *ate* requires an animate agent. An implausible sentence can be viewed as having both semantic and pragmatic violations; the sentence has a semantic violation because it violates selectional restrictions and it shows a pragmatic violation because the meaning is inconsistent with our event knowledge. Selectional restriction violation can determine the plausibility of a sentence. This is not to say that plausibility is a dichotomous category. Murray (2006) proposes the idea of conceptualizing plausibility as a continuum, where the plausibility of events is gradient, ranging from highly plausible to highly implausible.

Within the set of plausible events, a further distinction can be made based on the likelihood of occurrence of the events, that is, based on their typicality. Plausible

sentences can express an event that is more or less typical. Critically, implausible sentences have selectional restriction violations, while typical and atypical sentences pose no such violations. Making a distinction between typical and atypical events depends on their likelihood of occurrence, which needs to be assessed based on one's pragmatic knowledge of real world events.

Event typicality can also be conceptualized as a continuum, with events ranging from highly typical to highly atypical. Since all typical and atypical events are subsumed under the category of plausible events, the continuum for event typicality is essentially a subset of the plausibility continuum. A schematic representation is presented in Figure 2.1 below:

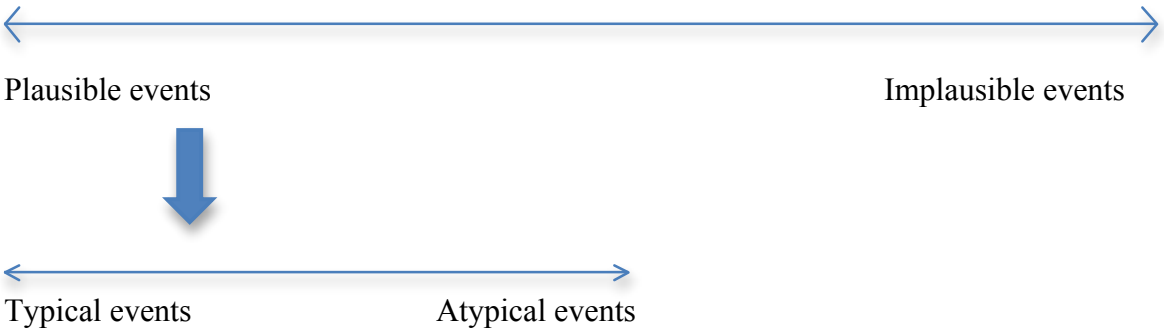

Figure 2.1: A schematic representation of the relationship between event plausibility and event typicality.

There is evidence showing that not all plausible sentences are processed similarly, Matsuki et al. (2009) showed that a plausible sentence that expresses an atypical event, such as (1b), resulted in a slower reading time and a larger N400 at critical regions than a plausible and typical sentence, such as (1a).

(1) a. Donna used the hose to wash her filthy car (typical)
b. Donna used the hose to wash her filthy hair (atypical)

Neither sentence is anomalous because both express plausible events and there are no selectional restriction violations; both *car* and *hair* are good candidates for the verb *wash*, as they are both washable objects. Selectional restrictions alone cannot explain the processing difference between (1a) and (1b), instead the difference in processing was attributed to the difference in event typicality.

Another way in which plausibility has been measured is by cloze probability in sentence comprehension. Federmeier and Kutas (1999) in their ERP study showed that the target words *pines* and *tulips* elicited greater N400 effects than the highly expected word *palms* in (2).

(2) They wanted to make the hotel look more like a tropical resort. So along the driveway, they planted rows of *palms/pines/tulips.*

All three sentence-final words were placed in the same plausible sentence context and this wider context, namely the scenario of *a tropical resort*, lead to the expectation of a specific type of tree to fit with the verb *planted*. If the fit between *planted* and the candidates *palm, pines,* and *tulips* were evaluated for their grammatical fit, all three

words were grammatical; however real world knowledge would distinguish *palms* as the more probable and appropriate fit than *pines* or *tulips*. The goal of this study was not to assess event typicality, but to show that people anticipated features of concepts based on preceding context (tree-like things). The results suggest a word that is not a good fit is one that is an unexpected and less typical item.

Federmeier et al. (2007) further showed that even within an isolated sentence that lacked discourse context, individual words were judged for their fit with the sentence context. They measured the N400 amplitude for expected and unexpected continuations as in example (3). Expected words elicited greater N400 amplitude in weak constraint contexts relative to strong constraint contexts. Unexpected words such as *collection* elicited similar N400 amplitudes in both weak and strong contexts.

(3) a. He bought her a pearl necklace for her *birthday/collection.* (Strong constraint context)
b. He looked worried because he might have broken his *arm/collection.* (Weak constraint context)

The findings were used as evidence that the depicted event (*he bought her a pearl necklace* and *he looked worried*) activates semantic features that are relevant to the context. Although typicality was not a concept explored in this study, the relevancy of these features appeared to be based on typicality.

Real world knowledge is however malleable, such that an implausible event that may appear to contradict real world situations can be overruled if it is situated in an appropriate context (e.g., Nieuwland & Van Berkum, 2006; Filik & Leuthold, 2008).

Nieuwland and Van Berkum showed in their ERP studies that given an appropriate context, the higher processing costs incurred by implausible real world sentences could be mitigated. They examined semantically anomalous sentences such as *the girl comforted the clock* and found a diminished N400 effect when it was preceded by a story-like context that anthropomorphized the agent noun: *the clock felt depressed.* In a second experiment, they created fictional contexts; two sentences were placed in the same discourse context where a peanut was described to be singing and dancing and obviously in a joyous mood. After reading this fictional context, the target sentence *the peanut was in love,* which on its own is regarded as anomalous, did not elicit a N400 effect. The plausible counterpart *the peanut was salted,* which according to our real world knowledge is plausible, was incompatible with the context and elicited a greater N400 effect. The results indicated that two words that are unassociated can be interpreted as acceptable given the appropriate context. This suggests then, the plausibility of an event is not just the association between two words, but the lexical association between two words can be overridden by context.

In a similar study, Filik and Leuthold (2008) also investigated whether implausible sentences could be accepted if they were placed in well-known story contexts such as The Hulk and Tom and Jerry cartoons. Their rationale for choosing cartoons was that the stories they chose had well-known fictional contexts. This is in contrast to Nieuwland and Van Berkum (2006), where participants may be unfamiliar with the given context. Filik and Leuthold suggested that an unfamiliar context might require participants additional time to contextualize the stories, potentially incurring processing

cost. In their study, Filik and Leuthold found that the target sentence *he picked up the lorry and carried on down the road* elicited an N400 effect when it was preceded by the sentence *Terry was very annoyed at the traffic jam on his way to work.* Real world knowledge dictates that it is highly unlikely for a man to pick up a truck, but when the protagonist was changed from *Terry* to *The Incredible Hulk*, the N400 effect diminished. Filik (2008) replicated this study using eye-tracking method and the results were consistent with the ERP results. Their findings provided further evidence that context can override the association between words and words that may not be associated can be placed in a sentence context that results in a plausible event.

In summary, the studies discussed in this section examined plausible sentences. While much is known how linguistic units are combined to form sentences that express plausible or implausible events, less is known about how a sentence is judged for its event typicality. The presence of selectional restriction violations can distinguish between plausible and implausible sentences. It is more challenging to use the presence of selectional restriction violation to establish a distinction between typical and atypical events, because atypical sentences usually denote events that are less likely to occur and may not have such violations. In theses studies, a violation occurred when the target word was not a good fit with the sentence or discourse context. This suggests then that the fit between words determines plausibility or typicality. The next section examines the possible ways that words are related or associated.

2.3 WORD ASSOCIATIONS

The plausibility or typicality of a particular event is not determined by individual words; rather a sentence formed by a group of words can be evaluated for event typicality (Bicknell et al., 2010). Words can be related in our memory network in different ways, by means of frequency of co-occurrence, meaning relation, grammatical requirements, or orthographic similarity. This section will examine some of the possible ways that words are associated with each other, in order to determine how the relationship between two words can contribute to the judgment of event typicality.

2.3.1 Noun associations

There are many ways in which nouns can be related or associated. Nouns that have similar orthographic forms have a neighborhood effect, whereby a word is processed faster in naming and lexical decision tasks if it has more orthographic neighbors (Andrews, 1997). Some studies found that participants were faster in making a lexical decision about the target word when it bore morphological similarity (words that were a regularly inflected variant of the prime word) with the prime word (e.g., Drews & Zwitserlood, 1995).

Aside from bearing physical similarity, words may be related in meaning. Words can be related or associated by their meaning, or by some form of semantic relationship. The general consensus is that words that are semantically related or associated have priming effects in naming and lexical decision tasks, whereby people are often faster to

decide whether the word is a real word or non-word if the prime and target words are semantically related (Hutchison, 2003).

Fischler's study (1977) showed that, in a lexical decision task, people were faster to make a lexical decision about the target word if the prime and target words were related or associated, compared to unrelated target-prime words. The word pairs they used were categorized into two groups, based on whether they were semantically related or associated. The criteria for relatedness were based on shared semantic features, for example, the word pair *bread-cake* is a related pair because both words have the feature [+baked goods], the word pair *boy-prince* is related because both words share the [+male] feature, and *team-staff* is related because both words denote a group of people. On the other hand, the criteria for word pairs that are associated appear to allow a wider range of types of relationships. A word pair can be associated because they are close synonyms, such as *road-street,* words that belong to the same taxonomy classification, *cat-dog,* words that usually co-occur together, *jump-rope,* or complementary pairs, *lock-key*. Similarly, in another study, also using a lexical decision task, Pecher et al. (1998) found priming effect for associated noun-noun pairs; their semantically associated pairs also consisted of a variety of noun pairs. These pairs included synonyms, *pretty-beautiful,* antonyms, *start-end, inside-outside,* and taxonomically related word pairs, *pig-boar*. These priming studies are informative in showing that two words that are related in meaning are processed faster, but the definition of 'relatedness' and 'association' are broad.

The nature of the semantic relatedness or association between noun pairs has been argued to be based on event knowledge. In the study by Hare et al. (2009), they examined whether priming could be obtained in various types of noun pairs. Nouns of different thematic classes were paired with another noun from another thematic class. The noun pairs reflected a real world knowledge based relationship, in contrast to the noun pairs used in other studies (e.g., Fischler, 1977) that examined pairs of nouns belonging to the same category or nouns that are related functionally or orthographically. Various types of noun pairs were examined in the studies of Hare et al., such as, event-people or event-objects (*sale-shopper, sale-luggage*), location-people or location-objects (*hospital-doctor, barn-hay*), and instrument-object (*bowl-cereal*) or people-instrument (*chef-knife*). Priming was obtained in both directions for all the noun pairs with the exception of instrument-people; it was suggested that the tested instrument nouns do not have a specific group of people that are associated with them, whereas certain people, such as *chef,* can have specific type of instruments that they are associated with, e.g., *knife.* Hare et al. argued that these word pair associations are driven by real world knowledge. The fact that we know certain words are associated, such as *hospital* and *doctor,* is because we know they are related based on what we know about the type of people who work at hospitals and the place where doctors typically work. This study demonstrated that people form an association between nouns based on their meanings. However it remains unclear whether these nouns that vary in their strength of association contribute to event typicality judgment and whether this information is used in real-time sentence processing.

2.3.2 Noun-verb associations

Fewer studies have examined the associative relationship between a noun and a verb. Ferretti et al. (2001) examined priming from verb to agent (*entertaining-comedian*), verb to patient (*entertaining-audience*), verb to instrument (*stirred-spoon*), and verb to location (*cooked-kitchen*). Priming effects were obtained in all prime and target pairs with the exception of verb and location. In a related study, McRae et al. (2005) examined whether priming could be obtained in noun-verb pairs, that is, whether nouns generate expectancies for a certain set of verbs. They examined agent-verb pairs (*actor-performing*), instrument-verb pairs (*axe-chopping*), patient-verb pairs (*ball-thrown*), and location-verb pairs (*airport-flying*). Robust priming was found when the primed verb was associated with the prime noun, and not when the prime noun was unassociated with the target verb. These priming studies suggest that words that are associated by means of real world knowledge are processed faster than words that are not associated in this way.

Interpreting the meaning of a sentence involves more than just analyzing the relationship between two words. In the next section, I will look at some of the studies that have examined how the relationship between two words can contribute to the interpretation of a sentence's meaning, and how the relationship between two words can potentially be changed by sentence context.

2.3.3 Combinatory effect of nouns and verbs

Verbs are governed by selectional restrictions that dictate the arguments they can select. The predictive power of verbs is evident in Altmann and Kamide's (1999) study, whose results showed that participants were more likely to look towards a picture of a cake when they heard *the boy will eat,* compared to when they heard *the boy will move.* Participants' eye movements were most likely guided by the knowledge of a verb's argument structure. Although *cake* is a possible patient noun for both *eat* and *move*, the probability that it occurs with *eat* is higher than with *move,* since *eat* requires an edible item and *move* requires a movable item that is not necessarily edible. It is evident then, that we cannot ignore the syntactic information that plays a role in assessing event typicality, but there are cases where event knowledge appears to be needed in addition to linguistic knowledge.

In the studies reported in Kamide, Altmann, & Haywood (2003), they showed that agent and verb combinations constrain expectations of the ensuing patient, and that this cannot be explained by lexical selectional information from the verb alone. In an eye-tracking study, participants were more likely to look toward a picture of a motorbike when they heard *the man will ride,* and toward a picture of a carousel when they heard *the girl will ride.* Both *motorbike* and *carousel* are appropriate patient arguments of the verb *ride,* but the lexical structure of the verb cannot explain why *motorbike* was preferred over *carousel* for the stimulus *the man will ride,* as neither noun violates selectional restrictions. The semantic feature of animacy does not provide sufficient explanation in this case, because animacy only indicates that the patient of *ride* must be a non-human entity. Both *motorbike* and *carousel* are non-human entities appropriate for

filling the patient role; the thematic fit of the patient noun is dependent on the event expressed by the sentence, and also the person who performs the action. Assessing the thematic fit between a verb and its arguments may require additional real world knowledge.

Bicknell et al. (2010) examined the interpretation of the object noun based on its congruency with the subject-verb pair. They conducted ERP and self-paced reading studies on sentences such as those in (4).

(4) a. The journalist checked the *spelling/brakes*.
b. The mechanic checked the *spelling/brakes.*

They found that while the sentences have no selectional restriction violations, the events expressed in these sentences vary in their likelihood of occurrence. The act of *checked the spelling* was seen as more congruent with *journalist* rather than *mechanic,* as evident in the higher processing cost when the subject *of checked the spelling* was *mechanic.* This was also true for *checked the brakes,* where there was a N400 effect and longer reading time at the patient noun when the agent was *journalist* instead of *mechanic.* The typicality of the action of *checked the spelling* or *brakes* is dependent on the identity of the person performing the action. The findings from this study showed that the combinatory effect of a verb and the agent determines the appropriate patient noun, and assessment of the typicality of the action is dependent on the identity of the agent. The studies also showed that sentences that have varying degrees of typicality are processed differently, with an atypical event likely to incur a higher processing cost than a sentence that expresses a typical event.

In a similar study, Matsuki et al. (2011), which was discussed in an earlier section, examined the fit of a verb and its agent, patient, and instrument arguments, such as *Donna used the hose to wash her filthy car/hair,* where *hair* was seen as the unlikely patient noun and therefore resulted in longer eye fixation times and reading times compared to the more congruent patient noun *car*. Matsuki et al. showed that the interpretation of an instrument-verb pair could be changed by the sentence context, and Bicknell et al. showed that the interpretation of a verb-patient pair was affected by sentence context, depending on the identity of the person initiating the action. In summary, the studies discussed in this section showed that people make judgments about event plausibility and typicality, but whether what has been reported is simply due to the lexical association between words cannot yet be excluded. The next section will look at studies that suggest the association between words could incur processing cost in a sentence.

2.4 WORD ASSOCIATIONS WITHIN A SENTENCE

Several studies show that context cannot overrule the semantic or pragmatic violations within a sentence. Kuperberg et al. (2007) examined the thematic fit between a verb and its arguments. They found that sentences in which there was a pragmatic violation within a sentence, such as *for breakfast, the boys would plant flowers in the garden* elicited a N400 effect because the verb *plant* is incompatible with the context of eating breakfast. In contrast, when there was a semantic violation in thematic fit, such as *for breakfast the eggs would eat toast and jam* and *for breakfast the eggs would plant*

flowers in the garden, there was no N400 effect but instead a P600 effect. Although *egg* is a relevant word in the *breakfast* scenario, it violated the animate agent requirement of *eat*. A P600 effect is an event-related potential where a positive deflection occurs around 500 milliseconds after the onset of the stimuli and 600 milliseconds is approximately the midpoint of this positivity after the onset (Coulson et al., 1998). A syntactic anomaly could be in the form of an ambiguous syntactic structure such as garden path sentences (Osterhout & Holcomb, 1992) or grammatical violation, specifically, subject-verb agreement, reflexive and number, and reflexive and gender agreement (Kaan et al., 2000; Osterhout & Mobley, 1995). Even in sentences with no structural or grammatical violations, there could still be a presence of P600 effect when the animacy requirement was violated, e.g., *at breakfast the eggs would eat everyday* (Kuperberg et al., 2003). The results reported by Kuperberg et al. (2007) suggest that although a word such as *egg* may be relevant in the context, it needs to be evaluated for its thematic fit with the verb; and a when there is a semantic violation of the fit between the agent and verb, there will be additional processing cost.

Hoeks et al. (2004) examined whether the local lexical-semantic relationship of words or the overall sentence meaning could predict the sentence-final verb in Dutch. They investigated the interaction between sentence-level constraints and local lexical-semantic fit between a verb and its arguments. The sentence-level fit was manipulated so that it was either a strong constraint for predicting the sentence-final verb, as indicated by the passive sentence structure *de speer werd door de atleten…* 'the javelin was by the athletes…' (which strongly predicted the verb to be *thrown*) compared to an active

sentence that was weak in constraining the upcoming sentence-final verb *de speer heeft de atleten*... 'the javelin has the athletes…' (which did not predict any particular verb). The lexical-semantic fit was manipulated so that a verb was either a good fit *geworpen* 'thrown' or a poor fit *opgesomd* 'summarized' with the preceding words, namely *javelin* and *athlete*. They found that there was an interaction between sentence-level constraint and lexical-semantic fit. When the verb was a poor fit, there was a significant difference in N400 amplitude compared to a verb that was a good fit. When the poor-fit verb occurred in a strong constraining context, the amplitude was significantly greater than that of the same verb in a weak constraining context. When the verb was a good fit, there was an absence of N400 effect regardless of the sentence context. This was unexpected given that in the weak constraining context, the sentence was anomalous, e.g., *the javelin has the athletes thrown,* although there was a P600 effect, a detection of syntactic anomaly. This study showed that a good lexical-semantic fit between content words does not incur processing cost, despite some of the stimuli sentences were semantically anomalous.

The issue that remains unclear is whether the association between two nouns can be judged for their typicality and whether this typicality can contribute to judgments of event typicality. Priming studies have shown that nouns that are associated can prime each other, whether they are related through a similar orthographic form or in some semantic way, and should be processed faster than noun pairs that are not related. We do not yet fully know whether the association between two nouns alone can contribute to the judgments of a sentence's plausibility or typicality. Furthermore, in the studies of

Bicknell et al., and Matsuki et al., they showed that the association between a verb and a noun contributes to the assessment of event typicality. The verb-noun pair *checked-spelling* can be seen as a typical action when combined with the agent *journalist,* but when combined with *mechanic,* the action becomes a less expected one. The typicality of the action of *checked the spelling* is therefore not absolute and is dependent on the person performing the action. If the association of a verb and noun pair can affect the typicality of an event in a sentence, there is reason to speculate whether this would also be the same for noun-noun pairs, since people form association between nouns based on real world knowledge, linguistic knowledge, or both.

There are some studies that examined the relationship between noun pairs and how the relationship between two nouns can affect event typicality. In the third experiment of Kamide, Altmann, & Haywood (2003), they examined whether the relationship between the agent-patient pair could generate expectations for the upcoming verb in Japanese, a verb-final language with case marking. They found that when presented with a sequence of two nouns, participants were more likely to look towards a theme object if the second noun was marked with the dative case compared to when the noun was marked with the accusative case. The crucial finding from this experiment was that, noun pairs can guide interpretation, before encountering the sentence-final verb. The relationship between two nouns can be inferred based on case markers in case-marking languages.

In a similar eye-tracking study, Kamide, Scheepers, & Altmann (2003) investigated whether case-marked nouns in German, a verb-final language, can predict

the upcoming noun. They found that participants were more likely to look towards a different target object depending on whether the first noun in the sentence was marked with nominative case or accusative case. For example, when participants heard *Der Hase frißt gleich* "the hare will shortly eat," when the noun was marked with nominative case, participants were more likely to look towards a picture of a cabbage since the nominative case indicates that the noun is an agent. When the same noun was marked with the accusative case (*Den Hasen*), this indicated that *the hare* is the theme of *eat,* and participants were more likely to look towards the picture of a fox even though the picture of a cabbage was also present. These studies showed that the participants used case information from nouns to anticipate upcoming input, before the verb appeared. In both studies of Kamide et al., case-markers guided the parsing and interpretation of sentence meanings. The relationship between the two nouns was indicated by case markers. However, case-markers are not available in every language.

The study of Pado and Zarcone (2011) provides insight into the contribution of noun pairs in sentence comprehension. They manipulated the association of two nouns, based on whether they are likely to occur together in the real world, for example, *baker-icing.* In their study, they examined logical metonymies in German, and manipulated the association of the verb with the preceding agent-patient pair. Metonymies were used because the interpretation of metonymies involves understanding an event that is implied but not explicit in the input, for example, *John began the book* implies that John began to read the book. They examined reading times for sentences containing a metonymic verb, an agent, patient, and action. Typical items such as *Das Kind-begann-Glasur-essen*

'child-began-icing-eat' and *Der Konditor-begann-Glasur-auftragen* 'baker-began-icing-spread;' were considered more typical than items with the combinations: *Das Kind-begann-Glasur-auftragen* 'child-began-icing-spread' and *Der Konditor-begann-Glasur-essen* 'baker-began-icing-eat'. The authors suggested that, the metonymic verb *began*, together with the agent and the patient nouns, generate expectations about the upcoming verb. The likelihood of the action being performed was dependent on the person performing the action and the object being acted on. They found that atypical sentences had longer reading times at regions after the target verb "eat/spread." They concluded that pragmatic real world knowledge determined the interpretation of logical metonymies, based on the reading time difference at regions after the verb between typical and atypical sentences. The studies discussed in this section suggest that the compatibility between a verb and its arguments, as well as the relationship between two nouns in a sentence, contributes to event plausibility but it is unclear whether the association between words, specifically noun pairs, contribute to event typicality judgment.

2.5 QUESTIONS TO BE ADDRESSED

Although it is clear that people make judgments on the typicality of events, it is still not fully clear how this judgment is derived. One of the factors that have not been examined fully is the relationship between two words, specifically, two nouns. Whether the association of these word pairs contributes to event typicality is not fully clear.

To address one of the main research questions of whether event typicality is more than just the semantic association between words, plausible sentences with varying degrees of typicality were examined in this dissertation. Attention was given in constructing the stimulus sentences to ensure that they described plausible events and contained no selectional restriction violations. The association of the noun pairs was manipulated and these noun pairs were then embedded into sentences that described events that varied in their typicality. Manipulating the association of noun pairs allows us to determine whether this association contributes to event typicality. The nouns were then paired with a verb that forms a sentence with SOV word order. SOV sentences allow the contribution of noun association to be examined in real-time sentence processing. The verb was also manipulated, it was either a typical or atypical action performed by the agent on the object. If the manipulation of the verb did not affect the assessment of event typicality, this shows that lexical association can override event typicality. On the other hand, if a manipulation of the verb affects the interpretation of event typicality, this suggests event typicality can override the association between the lexical association between two nouns.

The association between two nouns will need to be considered when processing the subsequent verb. Nouns that are related have a facilitation effect and they are processed faster than nouns that are not related. As seen in the study of Nieuwland and Van Berkum (2006), sentence context was able to override the associated between two unassociated words, although in Fischler's (1997) study, the association between two strongly associated words cannot be overridden, at least not initially.

2.6 Relevant typological characteristics of Chinese sentences

The studies conducted for this dissertation used Mandarin Chinese word combinations and sentences to take advantage of its simple morphological structure. Morphological markers such as past tense markers in English could contribute to a higher processing cost. The structural complexity of English verbs, which has various tense inflectional markers, has been considered to be the reason for their higher processing cost relative to the simple structure of nouns, which are only inflected for plurality (Tyler et al., 2004). In Chinese, nouns and verbs have equally simple morphological structures. Chinese, therefore, offers an opportunity whereby event knowledge integration can be evaluated without the concern that the structural complexity of nouns or verbs could incur different processing costs.

The canonical word order of Chinese is SVO and word order has been argued as an important strategy for sentence interpretation in Chinese (Li et al., 1993). The suggestion that word order is an important strategy in Chinese sentence processing is due to the limited use of inflectional markers in the morphology of Chinese, where there are no case or agreement markers that mark person, number, gender or tense. It would seem that attention to word order is therefore crucial for sentence processing. However, Chinese surprisingly allows many word order variations, such as OSV, SOV, and topic constructions. In general, languages that allow scrambling and free word orders have a rich case-marking system to indicate grammatical roles, but this is not the case for

Chinese. Furthermore, it is not uncommon to encounter sentences with null arguments since subject and object omissions are frequent in Chinese (Li & Thompson, 1981). The variations in word order are illustrated in the examples (5) – (8) below.

(5) wo chidiao pinguo le
1sg eat-finish apple PERF[4]
'I ate the apple.'

(6) wo ba pingguo chidiao le
1sg BA apple eat-finish PERF
'I ate the apple.'

(7) pingguo bei wo chidiao le
apple BEI 1sg eat-finish PERF
'The apple was eaten by me.'

(8) pinguo$_i$ wo chidiao $\varnothing_i$ le
apple 1sg eat-finish PERF
'As for the apple, I ate (it)'

The most common word order in Chinese is subject-verb-object (SVO) as shown in (5). The meaning conveyed by a SVO word order sentence can also be expressed using the *ba* construction, which has SOV word order, as shown in (6). Although (5) and (6) depict the same event, the use of the *ba* construction is considered the most appropriate when the expressed event implies affectedness, where an entity has undergone a change of state (Chu & Chi, 2006). The sentence in (7) shows another possible word order in Chinese, the passive *bei* construction, which has OSV word order. The sentence in (8) has the same words as that of a SVO sentence in (5) but the logical object *pingguo*

[4] List of abbreviations: BA = the *ba* marker, BEI = the passive marker *bei,* PERF = perfective marker, 1sg = first person singular

'apple' of the verb *chidiao* 'eat-finish' has been preposed to sentence-initial position. Topicalized constructions such as that in (8) are common in Chinese, which have lead linguists to consider Chinese as a 'topic-oriented' or 'discourse-oriented' language (Chao, 1968; Chu & Chi, 2006; Li & Thompson, 1981).

In addition, the selection of the initial noun phrase in Chinese is not always syntactically constrained; it is pragmatically driven and is tied to the predicate by the vague notion of 'semantic relatedness' (Li & Thompson, 1981). As seen in the examples below, the initial noun in (9) is neither a logical subject nor object of the verb; and the initial noun in (10) is not a logical subject of the verb. The notion of 'semantic relatedness' is frequently used to characterize the relationship between predicates, and reflects the non-rigidness of syntactic linear order.

(9) shuiguo wo zui xihuan pingguo
fruit 1sg most like apple
'(Among) fruit, I like apples best.'

(10) fan zhu-hao le
rice cook-finish PERF
'(somebody) finished cooking (a meal).'

The 'discourse-oriented' feature of Chinese makes it a good candidate for examining semantic processing. Another advantage of using Chinese sentences is the relatively free word order. The availability of the various word orders allows the contribution of individual words to event typicality in real-time sentence processing to be

examined. The same meaning can be expressed with sentences with SOV and SVO word orders without changing the meaning of the sentence.

2.7 SUMMARY

In summary, this chapter provided an overview of how people use their real world knowledge and linguistic knowledge to process sentences that vary in their plausibility and typicality. A sentence can express an event that is plausible; however, if there is a violation of selectional restrictions, the sentence is semantically anomalous and the event is implausible. Within plausible sentences, the likelihood of occurrence of the described events is different, thus a plausible sentence can express an event that is typical or atypical. The typicality of an event can be conceptualized on a scale. However, it is still not fully clear how the interpretation of event typicality is derived. Specifically, the association of two words in event typicality assessment will be the focus of this study. This dissertation will provide results from behavioral experiments that measured judgments of event typicality as well as examining how the judgments of word association and event typicality modulate real-time sentence comprehension.

The next two chapters report two sets of studies: a set of four offline typicality judgment studies and two online word-by-word self-paced reading studies. The judgment studies provide quantifiable measurements of word pairs and events that differ in their degree of typicality. These offline studies were used to obtain a measurement of the relationship between words based on how people judge their typicality of co-occurrence. The self-paced reading studies reported in chapter four examined how the association of

word pairs modulate sentence comprehension in real-time. The self-paced reading studies manipulated the association between two nouns and also the compatibility of a subsequent verb with the noun pair. These manipulations allowed the effect of the typicality of two nouns in sentence processing to be examined, as well as whether they contribute to a sentence's event typicality.

Chapter 3 Typicality judgment studies

In order to answer the question of whether event typicality is more than the semantic association between words, four offline judgment studies were conducted to obtain association norms and judgments of event typicality. These studies elicited native Chinese speakers' judgments on the strength of word associations and the typicality of the event expressed by a sentence.

Three factors were considered when conducting the typicality judgment studies: the association of two words, the word class of the word pairs, and the word order of the sentences. Two words could be associated in meaning; this association could be based on real world knowledge, linguistic knowledge, or both. Some words are strongly associated and some are weakly associated with each other. Creating word pairs of different word class combinations, such as subject-object and subject-verb, allows us to see if these various word pairs are judged differently for their association and how they contribute to event typicality judgment. Word pairs were embedded in sentences to evaluate the contribution of word pair association to judgments of event typicality. The word order of sentences was manipulated to determine whether event typicality judgment is affected by word order.

There are no known association norms available for Mandarin Chinese, and as such, it was necessary to obtain them. The first set of norms was obtained for subject-object pairs. These subject-object pairs were then used to create SOV sentences

expressing events that varied in typicality. The SOV sentences were than used as stimulus items in an online self-paced reading study, which will be reported in the next chapter. The study showed that the effect of verb association did not contribute to additional processing cost. In order to further determine the contribution of verbs in real-time sentence processing, two additional offline judgment studies were conducted: a subject-verb association judgment where participants judged the association of a subject noun and a verb, and SVO sentence judgment where participants judged the typicality of the event expressed by the sentence. The studies reported here will follow the order in which they were conducted, and their overall significance will be summarized at the end of the chapter.

A verb-object judgment study was also conducted but not reported in this chapter. Verb-object pairs could be judged for their association or events since Chinese allows subject omission. As a direction for future work, judgments of verb-object pairs were obtained but not included for further analysis. The results of the verb-object judgments are included in Appendix A.

Offline judgment studies provide quantitative measurements of typicality, and afford participants the opportunity to perform the task at hand without the pressure of time constraints (Schütze, 1996). The results from these offline studies then served as the basis for setting up self-paced reading studies to examine how the association of words affects event typicality interpretation in real-time sentence processing; these studies are reported in the next chapter.

3.1 SUBJECT-OBJECT (SO) ASSOCIATION JUDGMENT STUDY

The first typicality judgment study was conducted to examine the relationship between two nouns by obtaining judgment scores on typical and atypical subject-object[5] noun pairs. Verbs may appear to have more constraining power because they have inherent selectional restrictions that dictate what types of arguments they can select to create a well-formed phrase. Nouns on the other hand, do not have such constraints in their lexical representation. As discussed in the previous chapter, nouns can be related in various ways, either through form or through meaning. It is possible to judge noun pairs based on their semantic association. The studies reported in Hare et al. (2009) showed that nouns that are related have a priming effect. In Hare et al., their basis for creating noun pairs was based on how likely they are to be associated in the real world, such as a people-event pair *judge–trial,* an event-people pair *sale–shopper* and a location-item pair *garage–car.* Their study provided evidence that supported that nouns that may not share semantic features can nonetheless have a strong association based on real world knowledge.

Subject-object noun pairs were used in this study to obtain judgment scores for typical and atypical noun pairs, particularly noun pairs in Chinese. This noun pair typicality judgment study examines subject-object noun pairs by varying the object nouns that are paired with the same subject noun. Subject and object noun pairs were chosen

[5] The two nouns will be referred to as subject-object noun pairs because the pairs will be integrated into sentences in the next study. The relationship between two nouns in isolation may not appear to be a subject-object one, but the two nouns were designed as a subject and object in a sentence and will be referred here as a subject-object noun pair.

because the next goal is to determine the contribution of these noun pairs to event typicality interpretation in a sentence context and subject and object nouns can be considered as agent and patient nouns respectively in a sentence. Noun pairs that vary in their strength of association were judged to determine whether people judge them to be different based on association.

3.1.1 Method

Materials

Each participant read 256 subject-object noun pairs. The first noun (N1) was always animate and the second noun (N2) was always inanimate. This corresponds to the typical animacy characteristics of agent and patient, respectively (Dowty, 1991).

The second noun was manipulated, based on how likely it might be associated with the first noun. This created two possible N1-N2 pairs: strongly associated or weakly associated. A total of 128 N1s were used and each was paired with two strongly associated N2s and two weakly associated N2s. All N2s were chosen from the SUBTLEX-CH database (Cai & Brysbaert, 2010), which has frequency scores for individual Chinese characters as well as multi-character words. Frequency counts in this database were obtained by analyzing 46.8 million characters in subtitles from 6,243 Chinese films.

Noun pairs were created by varying N2 to create strongly or weakly associated noun pairs. Strongly associated noun pairs were created by identifying an N2 that is likely to be associated with a particular N1, based on the experimenter's intuitions. All

the weakly associated pairs were created by taking a N2 from a strongly associated noun pair and then paired that N2 with another N1; for example, *knife* is a strongly associated item for *chef,* but is weakly associated for *bus driver.* This was to ensure that the N2 items were used in both strongly and weakly associated noun pairs and that the weakly associated ratings for a noun pair were not caused by the infrequent occurrence or unfamiliarity of a particular N2. A sample stimulus set is presented in Table 3.1 below.

Strongly associated noun pairs	Strongly associated noun pairs
Example 1: *chushi–daozi* chef–knife	Example 1: *gongchesiji–daozi* bus.driver–knife
Example 2: *chushi–shipu* chef–cookbook	Example 2: *gongchesiji–shipu* bus.driver–cookbook

Table 3.1: Sample stimulus set for subject-object typicality judgment study.

A total of 512 items were created with an equal number of strongly associated and weakly associated noun pairs. The items were distributed into two lists and each list consisted of 256 items with an equal number of strongly associated and weakly associated noun pairs. Participants saw every N1 only once, which was paired with either a strongly associated N2 or an weakly associated N2. Participants were randomly assigned to one of the two lists and all the items were randomized for each participant.

Frequency scores for N2 items were obtained from the SUBTLEX-CH database and calculated to ensure there was no significant difference between the frequency of the stimulus items in the two lists. A Welch's two sample t-test was conducted to analyze the raw frequency of the stimulus items between the two lists. There was no evidence that the frequency scores between the two lists were statistically different, $t(251) = .26$, $p = .79$.

Procedure

All the studies reported in this chapter have the same testing procedure and recruitment process. Participants were asked to provide their judgments of noun pairs in a web-based questionnaire created on Survey Gizmo. For the subject-object judgment study, each participant saw 256 noun pairs with equal numbers of strongly associated and weakly associated items. They were given six practice items that were similar to but not identical to any of the experimental items. Participants were asked to provide a judgment score on the noun pairs on a scale of 1 to 7, where 1 indicated that the nouns were weakly associated and 7 was strongly associated. Participants were asked to judge on the likelihood of association/relatedness of the two nouns based on their intuition.[6]

Participants

Participants were recruited from the experimenter's social network and the questionnaire link was posted on social media websites and re-posted by participants. All the participants were screened for their nationality and for whether their first language

[6] Instruction in Chinese: 請依據您的直覺反應，來評估兩者名詞關連性。

was Mandarin Chinese. As there are considerable dialectal differences between the different varieties of Mandarin Chinese, only speakers of the variety of Mandarin spoken in Taiwan were considered for this study. As it is common for Taiwanese citizens to be fully fluent in both Mandarin and the local dialect of Taiwanese, participants were asked to indicate the age at which they started learning Mandarin if they had learned Taiwanese as their first language. Data from participants that speak Taiwanese as their first language were included for analysis if they indicated they learned Mandarin during elementary school or earlier. Participation was restricted to those between the ages of 18 and 35. All the participants volunteered their time without compensation.

Sixty-eight (68) people participated in this subject-object judgment study (43 female, 25 male). Of the two lists, 33 participants were assigned to list 1 and 35 participants were assigned to list 2. The average age of the participants was 28.35 years (*SD* = 4.33, range = 18 to 35). Three participants indicated that they had learned Taiwanese as their first language and began learning Mandarin at the ages of 1, 3, and 8. Data from these three participants were included for analysis.

3.1.2 Results

The mean score for strongly associated items was 6.37 (*SD* = 1.11) and the mean score for weakly associated items was 1.70 (*SD* = 1.33). A 2x2 analysis of variance (ANOVA) analysis was conducted, with the association of the noun pairs as a within-subjects factor and item list as a between-subjects factor. The dependent variable in this study, as well as in all the studies reported in this chapter, was judgment score. The

between-subjects factor was included to ensure that the variation in judgment scores was not caused by the items in the two different lists. The ANOVA showed a main effect of noun association, $F(1, 65) = 3107.42$, $p < .001$, no main effect of list, $F(1, 65) = .20$, $p = .66$, and no interaction between noun association and list, $F(1, 65) = 3.05$, $p = .086$. The main effect of noun association indicates that the difference in scores between strongly associated and weakly associated items was significant. A non-significant main effect of item list indicates that there was no significant variation between the types of items across the two lists. The adjusted R-squared for this model was 0.79. Removing the list variable from the model, the adjusted R-squared remained the same. This value indicates noun association explains a large amount of the variance in judgment scores.

Results from this study revealed a systematic difference in participants' judgments of subject-object pairs. Noun pairs that were weakly associated had lower judgment scores than strongly associated noun pairs. The next judgment study will use the results from this study to examine whether the association of two nouns contributes to event typicality judgment when theses noun pairs are embedded in a sentence.

3.2 SUBJECT-OBJECT-VERB (SOV) TYPICALITY JUDGMENT STUDY

The subject-object judgment study showed that participants distinguished the association of the noun pairs based on how likely the nouns were to occur together. The next step in understanding the role of this associative relationship was to examine whether the association of a noun pair contributes to event typicality assessment. When

noun pairs were judged independently, the association of the noun pairs was determined by how likely they were related. When placing a noun pair in a sentence context, however, the event expressed by a sentence needed to be evaluated and the typicality of the noun pair might or might not influence the interpretation of the sentence's judged event typicality. This judgment study investigated how strongly associated versus weakly associated noun pairs contribute to event typicality judgments.

Stimulus items were created based on the results from the subject-object association judgment study. Strongly associated and weakly associated subject-object pairs were combined with a verb that was manipulated for its compatibility with the noun pair. A noun pair was combined with a verb that was a typical action or atypical action performed by the protagonist on the object item. The assessment of event typicality is the combination of the inanimate noun and the verb relative to the animate noun. For a strongly associated noun pair, pairing it with a typical verb creates a typical event and pairing it with an atypical verb creates a less typical event. For a weakly associated noun pair, the atypical judgment could potentially be reversed with a typical verb, or the atypical interpretation could be reinforced with an atypical verb. These manipulations created four unique conditions that reflected a range of event typicality scenarios. It was expected that the range of typicality scores would reflect this variability in event typicality.

3.2.1 Method

Materials

In the subject-object association judgment study, 128 N1s were created and each was paired with two strongly associated N2s and two weakly associated N2s. In this study, for each N1, one strongly associated N2 and one weakly associated N2 were selected to form a strongly associated and weakly associated noun pair.

The selection of an N2 was based on the scores from the subject-object association judgment study. To create a strongly associated noun pair, one of the two strongly associated N2s with a higher score was selected to form a pair with the N1. There were a few exceptions to this procedure for selecting strongly associated and weakly associated N2s. Given that the same N2s were used for the strongly associated and weakly associated items, the scores for the N2s had to be verified for both the strongly associated and weakly associated noun combinations to ensure a good fit. For example, the N1 *postman* was paired with the strongly associated N2s *package* and *letter* with similar mean scores of 6.94 and 7.00 respectively but when these same nouns were used as weakly associated N2s to pair up with the N1 *referee,* the scores for this weakly associated noun set were 1.41 (*referee–package*) and 1.94 (*referee–letter*). If only scores for the strongly associated items were considered, the item *letter* would have been chosen as the weakly associated item for *referee* even though *package* had a lower score. In cases such as this, if the strongly associated N2 pairs had very similar scores for a given N1, the scores from the weakly associated pair were considered in the selection process.

In the case of *package and letter, package* was preferred over *letter* because this created a better weakly associated item but did not have much effect on the strongly associated noun pair. A few cases where this rule did not apply were when it was decided based on intuition that it would be difficult to create a natural sounding sentence from the noun pairs.

After a unique strongly associated N2 and an weakly associated N2 were selected for the subject noun, the final average score for all the typical noun pairs was 6.54 and all the atypical pairs had a score of 1.54. Each of the 128 N1s was paired with a strongly associated N2 and n weakly associated N2. A total of 256 noun pairs were used in this study with equal numbers of strongly associated and weakly associated noun pairs.

Every strongly associated and weakly associated noun pair was assigned two verbs; a typical verb, which together with the noun pair formed a sentence that expressed a typical event and an atypical verb that created a less typical event. The verbs were selected based on the experimenter's intuitions. For example, the strongly associated noun pair *chef–knife* was paired with the verbs *wash* and *mail* denoting typical and atypical actions respectively, thereby creating typical and atypical events. These two verbs were then paired with the weakly associated counterpart of the noun pair, yielding four unique conditions; Table 3.2 below presents a sample stimulus set. As it is challenging to find verbs that can form a natural sounding combination with the noun pairs, a few verbs were used repeatedly, but this was minimized whenever possible.

		Noun typicality[7]	
		Typical	Atypical
Verb typicality	Typical	**Condition TNTV** *chushi–daozi–xiganjing* chef–knife–wash	**Condition ANTV** *chushi–micaifu–xiganjing* chef–military.uniform–wash
	Atypical	**Condition TNAV** *chushi–daozi–jichuqu* chef–knife–mail	**Condition ANAV** *chushi–micaifu–jichuqu* chef–military.uniform–mail

TNTV = typical object noun typical verb, ANTV = atypical object noun typical verb, TNAV = typical object noun atypical verb, ANAV = atypical object noun atypical verb

Table 3.2: Factors and levels for SOV typicality judgment study.

The crossing of the two factors of noun typicality and verb typicality resulted in four unique conditions. Therefore for each stimulus set, there were four different sentences reflecting the four typicality scenarios. A total of 128 stimulus sets were created with fours items in each set, which resulted in a total of 512 unique sentences. The noun and verb combinations were constructed as simple SOV sentences using the Mandarin *ba* construction. A sample sentence is presented in (1) below:

(1) chushi ba daozi xiganjing-le
chef BA knife wash-PERF
'The chef washed the knife.'

Four lists were created and the items were assigned to each list using a Latin Square design. Only one item from each stimulus set occurred in a given list. Each list

[7] Noun pair association will be termed 'noun typicality' in this study since the object nouns denotes either a typical or atypical object in the SOV sentences.

had 128 items and an equal number of items from each condition and equal number of typical and atypical items. Participants were randomly assigned to a list and items were randomized for each participant.

Procedure

The processes for recruiting participants and screening were identical to those of the subject-object judgment study.

Participants were asked to provide their intuitions on the typicality of the events described in the SOV sentences. Since there is no direct translation for the word 'typicality', participants were asked to judge the likelihood of the action being performed by the protagonist.[8] They were asked to rate each item on a scale of 1 to 7 with 1 being the least likely scenario and 7 being the most likely scenario. Sample items were given depicting scenarios with high and low likelihoods.[9]

Scores were then averaged for each item across participants and compared between different conditions to determine the typicality of the different types of sentences formed by the different combination of the nouns and verbs.

Participants

Fifty-eight (N = 58) people participated in this typicality judgment study (38 female, 20 male); 18 were assigned to List 1, 12 to List 2, 12 to List 3, and 16 to List 4.

[8] Instruction in Chinese: 請根據您的直覺，來評估句子描述裡，主角做這件事的可能性。

[9] For example, the sentence of 'the bird built a bird's nest' was given and this was noted as 'an event that is very likely to take place' while 'the sailor built a bird's nest' was provided as an example of 'an event that is very unlikely to take place.'

The average age of the participants was 28.10 years (*SD* = 4.25, range = 18 to 34 years). One participant indicated that he/she learned Taiwanese as their first language and began learning Mandarin at 3. Data from this participant was included for analysis. Participants volunteered their time without compensation.

3.2.2 Results and discussion

The design of the study was a three-way 2x2x4 ANOVA, with two within-subject factors of noun typicality and verb typicality. A between-subject factor of item list was again included to ensure there was no variation between the items among the lists. Mean scores and standard deviations for each condition are presented in the Table 3.3 below.

	Condition TNTV *chef–knife–wash*	Condition ANTV *chef–military.uniform–wash*	Condition TNAV *chef–knife–mail*	Condition ANAV *chef–military.uniform–mail*
Mean (SD)	6.45 (1.14)	2.81 (1.93)	4.58 (2.18)	2.42 (1.73)

Table 3.3: Mean scores and standard deviations for each condition in SOV typicality judgment study.

The ANOVA showed main effects of noun typicality, $F(1, 51) = 739.79$, $p < .001$, and verb typicality, $F(1, 51) = 401.46$, $p < .001$, as well as a significant interaction of these two predictors, $F(1, 51) = 175.94$, $p < .001$. Post-hoc analysis with Tukey's

correction showed a significant difference between each of the four combinations of levels for noun and verb typicality, all $p < .001$.

There was no main effect of list, $F(3, 51) = 0.70$, $p = .56$, no interaction between item list and noun typicality, $F(3, 51) = 1.13$, $p = 0.35$, and no interaction between item list and verb typicality, $F(3, 51) = 0.30$, $p = 0.82$. There was no interaction between all three predictors, $F(3, 51) = 1.43$, $p = .25$. The adjusted R-squared for this model was 0.450. Removing the list variable from the model, the adjusted R-squared for the model with noun and verb typicality was 0.445.

Sentences that had two typical items (typical object noun, typical verb), which depicted a highly typical event, as expected, had the highest mean score. Sentences with two atypical items (atypical noun, atypical verb) had the lowest mean score, which was also expected. Considering now the sentences that have only one typical item, sentences that had a typical noun and atypical verb (*chef–knife–mail*) had higher scores than sentences with an atypical noun and typical verb (*chef–military.uniform–wash*). In fact, *chef–military.uniform–wash* was rated almost as low as the sentences with two atypical items, which was an unexpected finding. The mean scores for *chef–military uniform–wash* and *chef–military.uniform–mail* were similar, although this difference was nonetheless significant with post-hoc analysis performed using Tukey's correction, $t(51) = -4.93$, $p < .001$.

The similarity between the scores for sentences with an atypical noun pair and a typical verb, and sentences with an atypical noun pair and an atypical verb suggests that the association between two nouns had significant contribution to event typicality. The

mean score for *chef–knife–mail* (TNAV) was closer to the most typical items and participants found the event expressed by these sentences to be more typical than those in condition *chef–military.uniform–wash* (ANTV). The event expressed by an ANTV sentence was considered as being unlikely although the manipulation was aimed at creating a typical event. This showed that the association of nouns contributes to event typicality and the weak association between two nouns could not be reversed by a typical action.

In order to confirm this observation, the next judgment study also examined the rating of sentences with two nouns and verbs, but the word order was now changed to subject-verb-object (SVO). The goal was to determine whether the effect of the noun association would be consistent when a verb appeared between the two nouns.

3.3 SUBJECT-VERB-OBJECT (SVO) TYPICALITY JUDGMENT STUDY

The SOV judgment study provided judgment scores on the four scenarios depicting different degrees of event typicality. In order to determine whether the judgment of typicality is affected by word order, a subject-verb-object typicality judgment study was conducted to determine if the results would be consistent with that of the SOV study.

The stimuli used for this study were identical as those used for the SOV typicality judgment study with the exception of a different word order. In the SOV study, the verb occurred after a sequence of two nouns and in this study, the verb was positioned

between the subject and object nouns. The manipulation of this study followed that of the SOV judgment study, where the association of the noun pair, as well as the typicality of the verb, were manipulated. The noun and verb combinations were constructed as simple Mandarin SVO sentences. A different word order does not change the meaning of the sentence and in a SVO sentence, a *ba* marker is not required, which was required in a SOV sentence. A sample SVO stimulus sentence is presented in (2) below:

(2) chushi xiganjing-le daozi
chef wash-PERF knife
'The chef washed the knife.'

A sample stimulus set is presented in Table 3.4 below.

		Noun pair typicality	
		Typical	Atypical
Verb typicality	Typical	**Condition TNTV** *chushi–xiganjing–daozi* chef–wash–knife	**Condition TNAV** *chushi–xiganjing–micaifu* chef–wash–military.uniform
	Atypical	**Condition ANTV** *chushi–jichuqu–daozi* chef–mail–knife	**Condition ANAV** *chushi–jichuqu–micaifu* chef–mail–military.uniform

TNTV = typical object noun typical verb, ANTV = atypical object noun typical verb, TNAV = typical object noun atypical verb, ANAV = atypical object noun atypical verb

Table 3.4: Factors and levels for SVO typicality judgment study.

A different word order may or may not affect the judgment of event typicality. If the association of the nouns contribute more significantly than the typicality of the verb to event typicality, the judgment of the events in this study will be similar to those in the SOV study. The canonical word order in Mandarin is SVO, although the *ba* construction used in the SOV judgment study is also frequently used. To ensure the findings in the SOV study were not caused by word order, this follow-up study with SVO studies was needed.

The materials used in this study were based on the items from the SOV judgment results. In the SOV study, participants were asked to provide judgment scores based on the typicality of the event expressed by the sentences. Because only 72 sets of stimulus sets were used in the online self-paced reading study reported in the next chapter, this study only included those 72 stimulus sets. These 72 sets were selected based on the distribution of the scores and reflect a diverse range of scores between the four conditions.

3.3.1 Method

Materials

Participants were asked to provide their intuitions on the event typicality expressed by SVO sentences. The experimental sentences were distributed into four lists, with 72 sentences in each list and an equal number of sentences from each condition. Only one sentence from each set appeared in a given list. Each participant saw 72 sentences and was asked to provide a judgment score on the typicality of the events on a

scale of 1 to 7. A score of 1 indicates that the sentence expresses a highly atypical event, and a score of 7 indicates that the event is highly typical.

Procedure

The processes for participants recruiting and screening were the same as the previous studies.

The instructions participants received for this study were identical to the SOV judgment study, since stimulus items had the same noun and verb combinations but in a different word order. Participants were asked to judge the likelihood of the events described by the SVO sentences, and rate the sentences on the same 1 to 7 scale. Scores were averaged for each item and compared between different conditions to determine the typicality of the different types of sentences formed by the various combinations of nouns and verbs.

Participants

The total number of participants for this typicality judgment study was 37. One participant rated all the sentences as "7" and the other participant was not of Taiwanese nationality. Data from these two participants were removed from consideration. The total number of participants whose data were included for analysis was 35 (21 female, 14 male). The mean age of the participants was 30.66 (*SD* = 3.37, range = 18-35). All participants indicated that they learned Mandarin as their first language. Participants volunteered their time without compensation.

3.3.2 Results and discussion

The design of this study was identical to that of the SOV study. This study also employed a three way 2x2x4 ANOVA design with two within-subjects factors of noun typicality and verb typicality. A between-subjects factor of item list was also included to ensure there was no variation between the items amongst the lists. The items were distributed into four lists using a Latin Square design and each list contained equal numbers of items from each condition. Participants were randomly assigned to one of the four lists. The mean scores and standard deviations for each condition are presented in Table 3.5 below.

	Condition TNTV *chef–wash–knife*	Condition ANTV *chef–wash–military.uniform*	Condition TNAV *chef–mail–knife*	Condition ANAV *chef–mail–military.uniform*
Mean (SD)	6.37 (1.24)	2.75 (1.84)	4.44 (2.16)	2.25 (1.55)

Table 3.5: Mean scores and standard deviations for each condition in the SVO typicality judgment study.

The ANOVA showed that there were main effects of noun typicality, $F(1, 30) = 412.11$, $p < .001$, and verb typicality $F(1, 30) = 163.37$, $p < .001$, as well as an interaction between these two, $F(1, 30) = 89.34$, $p < .001$. Post-hoc analysis with Tukey's correction showed a significant difference between each of the four combinations of levels for noun

and verb typicality, all $p < .001$. The between-subjects factor of item list was not significant, $F(3, 30) = 0.92$, $p = .44$.

The interaction between item list and noun typicality was not significant, $F(3, 30) = 0.77$, $p = .52$, but the interaction between item list and verb typicality was significant, $F(3, 30) = 4.64$, $p < .01$. There was no three-way interaction, $F(3, 30) = 0.26$, $p = .85$. The mean score for *chef–wash–military.uniform* and *chef–mail–military.uniform* were very similar and yet this difference was significant $t(30) = 4.01$, $p < .01$. The items were rated on a scale of 1 to 7 and a small difference might show up as statistically significant. The adjusted R-squared for this model was 0.47. Removing the list variable from the model, the adjusted R-squared for the model with noun and verb typicality was 0.46, which was similar to the adjusted R-squared of 0.45 for the equivalent model in the SOV typicality study.

In the SVO study, sentences that had only one atypical item had mean scores that were higher than sentences with two atypical items and lower scores than sentences with two typical items. However, a sentence with a typical noun and an atypical verb (*chef–mail–knife)* had higher mean scores than sentences with a typical verb and an atypical noun (*chef–wash–military.uniform).* This again suggests that the association of the noun pairs had more significant contribution to event typicality than verb typicality does and is not affected by word order.

The studies reported thus far suggested a robust effect of noun typicality. The effect of verb typicality was significant in both SOV and SVO judgment studies, and there was an interaction between noun typicality and verb typicality, where sentences that

had an atypical noun had lower mean scores regardless of the verb in the sentence. In order to assess the contribution of verbs in event typicality assessment, a subject-verb association judgment study was conducted to determine the association between a subject noun and a verb. In a subject-verb association judgment study, the only manipulation was the association of the verb with the subject noun, which allows the contribution of verbs to be examined, without the influence of the object noun.

3.4 SUBJECT-VERB (SV) ASSOCIATION JUDGMENT STUDY

The SOV and SVO judgment studies showed that the scores across the four different typicality conditions were significantly different from each other. So far the studies have shown that the association of noun pairs contributed more to event typicality judgment than verb typicality did. In both SOV and SVO studies, verbs contributed to event typicality in a sentence but it is not fully clear whether verbs alone could contribute to event typicality.

A subject-verb association judgment study was conducted to isolate the contribution of verbs to judgments of event typicality. The question of interest was, in the absence of a post-verbal object, would we be able to see a differential judgment on the subject-verb pairs? Without an object noun, how will people evaluate the association between a subject noun and a verb? Studies showed that nouns prime verbs (e.g., McRae et al., 2005), and the goal of this study was to determine whether verbs that are associated with certain actors differ in their judged association.

3.4.1 Method

Material

There were 72 sets of subject-verb stimuli sets and each set consisted of two subject-verb pairs. The association between the subject and the verb was manipulated, based on their likelihood of association. Every noun was paired with a typical and an atypical verb, denoting a typical and atypical action performed by the protagonist. These item pairs were identical to the stimulus items of the SVO judgment study, but with the object noun omitted. A total of 144 items were used as stimuli with an equal number of typical and atypical items. Two lists were created and items were distributed amongst the two lists and each list consisted of 72 unique subject nouns and half of the subject nouns were paired with a typical verb and half with an atypical verb. Participants were randomly assigned to a list and item presentation was randomized for each participant.

Procedure

The processes for recruiting participants and screening were identical to that of the previous studies. In order to facilitate participant recruitment, the age range of the participants was extended from between 18 and 35 to 18 and 45.

Participants were asked to provide a judgment score on the subject-verb pairs using a 1 to 7 scale. Participants were instructed to assess the likelihood of the protagonist performing a given action[10], with a score of 1 indicating the action denoted

[10] Instruction in Chinese: 請根據您的直覺，來評估句子描述裡，主角做這個動作的可能性。

by the verb is highly unlikely to be performed by the actor while a score of 7 indicates that the action is highly likely to be performed by the actor.

Participants

The total number of participants for this study was 51 (28 female, 23 male). There were 29 participants assigned to list 1 and 22 participants assigned to list 2. The average age of the participants was 30.76 (*SD* = 5.10, range = 19-42). All participants reported that they are native speakers of Mandarin Chinese from Taiwan.

3.4.2 Results and discussion

This study employed a 2x2 ANOVA design with a within-subjects factor of verb typicality and a between-subjects factor of item list. The mean score for the typical subject-verb pairs was 5.34 (*SD* = 1.76) and 3.76 (*SD* = 1.88) for the atypical pairs. The ANOVA showed that there was a main effect of verb typicality, $F(1, 49) = 498.07$, $p < .001$ and no main effect of item list, $F(1, 49) = 0.06$, $p = .80$. There was no interaction between verb typicality and item list, $F(1, 49) = 0.53$, $p = .47$. The R-squared for this model is 0.16. Removing the list variable from the model, the adjusted R-squared for the model with the variable verb typicality remains the same, 0.16.

The results suggest that the typicality of the verb does affect the association judgment of subject-verb pairs; people do make a differentiation as to the type of actions that are more likely to be performed by a given agent. However, the effect of verb

typicality was not always apparent in judging event typicality in SOV and SVO sentences. In those studies, the typicality of the event denoted by the sentences had to be evaluated, and this process involves integrating the contribution of both nouns and verbs.

3.5 DISCUSSION AND SUMMARY

The four offline judgment studies reported in this chapter suggested that event typicality is not just simply the semantic association between words. The association between noun pairs and subject-verb pairs both contribute to the sentence's event typicality. Participants were sensitive to the association of subject-object and subject-verb pairs, such that strongly associated pairs had higher mean scores than weakly associated pairs, as shown in Table 3.7[11] below. However, comparing subject-object and subject-verb associations, the association of nouns showed greater distance in scores between strongly associated and weakly associated pairs than the contrast of subject-verb pairs. Although the association of subject-verb pairs was not as strong as subject-object pairs, the difference between a typical and atypical verb was nonetheless significant.

[11] The mean and standard deviation for subject-object pairs were calculated for the 72 sets of items that were used in the SVO and SV studies. This allows a direct comparison between the subject-object and subject-verb pairs.

	Strongly associated pairs	Weakly associated pairs
Subject-object Mean (SD)	Typical object *chef-knife* 6.43 (1.07)	Atypical object *chef-military.uniform* 1.66 (1.27)
Subject-verb Mean (SD)	Typical action *chef-wash* 5.34 (1.76)	Atypical action *chef-mail* 3.76 (1.88)

Table 3.7: Comparison of mean scores and standard deviations for each condition in subject-object and subject-verb association judgment studies.

When these subject-object and subject-verb pairs were embedded in SOV and SVO sentences, results from the SOV and SVO judgment studies showed that both nouns and verbs contribute to event typicality. The scores across conditions were similar in both studies, which are presented in Table 3.8[12] below.

	Condition TNTV *chef–knife–wash*	Condition ANTV *chef–military.uniform–wash*	Condition TNAV *chef–knife–mail*	Condition ANAV *chef–military.uniform–mail*
SOV Mean (SD)	6.51 (1.02)	2.84 (1.92)	4.44 (2.15)	2.29 (1.64)
SVO Mean (SD)	6.37 (1.24)	2.75 (1.84)	4.44 (2.16)	2.25 (1.55)

Table 3.8: Comparison of mean scores and standard deviations for each condition in SOV and SVO typicality judgment studies.

[12] The mean and standard deviation for the SOV sentences were calculated for the 72 sets of items that were used in the SVO and SV studies. This allows a direct comparison between SOV and SVO sentences.

The similar findings in the SOV and SVO judgment studies suggest that the effect of noun association was not modulated by the presence or absence of an intervening verb between the two nouns. Word order was not relevant in event typicality judgment, since the scores were almost identical in SOV and SVO studies.

Both noun and verb individually contribute to event typicality, but when they are combined together nouns had a more significant effect. The stronger effect of nouns was possibly due to the stronger association between subject and object nouns. This suggests the association of nouns in these stimulus sets has more influence than verbs in their contribution to event typicality. There was a sub-additive effect in event typicality. The difference in scores between the least typical condition ANAV and the most typical condition TNTV, was not a total of the difference between ANTV and TNTV, and TNAV and TNTV. Having an atypical object had a more significant effect than an atypical verb, as reflected in the lower scores in condition ANTV than condition TNAV. This is not to say that verbs are not contributing to event typicality. In the subject-verb judgment study where participants judged the association of word pairs consisting of a subject noun and verb, there was a main effect of verb typicality.

A word pair does not describe an event, certainly noun pairs are not events, although the subject-verb pairs used in the studies could potentially be construed as meaningful sentences. This is because Chinese allows pro-drop where subject and object omissions are prevalent (Li & Thompson, 1981). Regardless of whether the subject-verb pairs were judged for their association or for event typicality, the results were consistent.

Participants received different instructions in the studies based on whether they were judging word pair association or sentence event typicality. In the word pair judgment studies, participants judged the association between the words. When these word pairs were embedded in a sentence, the task was different. The instructions and examples explicitly guided them to think about the event denoted by the sentence, not just the association between words. Although the tasks for judging word pairs and sentences were different, nouns consistently contributed more to event typicality than verbs did.

To answer the next research question of whether typicality modulates sentence processing in real-time, the results obtained in the studies reported in this chapter formed the basis for designing the stimuli for two online sentence comprehension experiments. The next chapter reports the results of two self-paced reading studies that examine whether people use the information reported in this chapter – word association judgments and event typicality assessments – in real-time sentence processing.

Chapter 4 Typicality in real-time sentence comprehension

The results of the offline judgment studies presented in the previous chapter showed that nouns and verbs play significant but dissimilar roles in how we judge the association of subject-object and subject-verb combinations with respect to event typicality. The results suggest that the association between two nouns had a more significant effect on event typicality assessment than subject-verbs do. In order to determine whether these offline subjective judgments are used in real-time sentence processing, two self-paced readings were conducted. This chapter reports on two self-paced reading studies that assess whether offline findings also apply in real-time sentence processing.

Psycholinguists use self-paced reading studies to evaluate the incremental nature of language processing (Just et al., 1982). Reading times throughout the sentence at each individual word are highly sensitive to the processing demands at that moment and spillover effects from previous words. A slow down in reading time is indicative of processing difficulty in the linguistic stimuli, allowing us to localize the point at which the sentence incurs a higher processing load for the reader (Mitchell, 2004).

4.1 Experiment 1: SOV self-paced reading study

The purpose of the first self-paced reading experiment was to assess whether there is a difference in reading time at critical regions in sentences that express either a typical or an atypical event. The sentences used in this study have SOV word order. The typicality of the noun pair, as well as the association of the verb with the noun pair, were manipulated. The manipulations allowed us to examine whether the association of the two nouns and typicality of an action contribute to event typicality assessment. In this chapter, the typicality of the noun pair refers to the association of the two nouns, while the typicality of the verb refers to how likely it is to be associated with the noun pair.

The syntactic construction used in this study was the Chinese *ba* construction. The word order of the *ba* construction is SOV, which deviates from the canonical SVO word order. Agent-patient noun pairs are chosen because they typically denote a relationship where one entity is acting on another. Since the *ba* construction is frequently used in Chinese, there is no concern that speakers are unfamiliar with a non-canonical SOV construction. An example of a *ba* sentence is shown in (1).

(1) laoshu ba chisi chidiao le
mouse BA cheese eat PERF
'The mouse ate the cheese.'

The object noun in the *ba* construction is preceded by the marker *ba*, as in (1). In canonical *ba* sentences, the *ba* marks the direct object, which is the affected noun (Chu & Chi, 2006; Li & Thompson, 1981).[13]

SOV sentences allow the association between nouns to be examined and to see how it contributes to the typicality of the event expressed in a sentence. Although the word order of the *ba* construction deviates from the canonical SVO word order of Mandarin Chinese, the order of the agent and patient nouns remains the same as in the SVO sentences. Ferreira (2003) found that non-canonical sentences in English are difficult only if they deviate from this canonical agent-patient order; the complexity of the syntactic structure is not what caused the higher processing cost in non-canonical sentences. Thematic role assignment in *ba* sentences follows the same order as that of a canonical SVO sentence and therefore these sentences are not expected to incur additional processing costs for understanding agent-patient relationship.

In fact, there appears to be little, if any, additional processing cost associated with the *ba* construction (Philipp et al., 2008). In contrast, another non-canonical Chinese syntactic construction, the *bei* construction where the order is OSV, showed evidence of increased processing costs in Philipp et al.'s study. The same study showed no N400 effect at various points in the *ba* sentences. The effect of processing a non-canonical SOV construction is therefore not of concern here.

---

[13] There are cases of the *ba* construction where it is not clear whether the post-*ba* object noun is affected, such as (2). However, this kind of usage is infrequent and will not be considered.

(2) ta ba ni xiang de fan dou chi-bu-xia
3sg BA 2sg miss DE food even eat-no-swallow
'He misses you so much that he won't eat his meals.'

Using a sentence with SOV word order affords two points of interpretation for which typicality judgments can be assessed. The points of interest in a *ba* sentence are the position of the post-verbal noun (N2) and the following regions, as well as the verb and its subsequent regions. N2 is strongly or weakly associated with N1. A strongly associated N1-N2 pair indicates that the combination of N1 and N2 is a typical and expected one, for example *baby–milk bottle.* A weakly associated N1-N2 pair is one where the combination of N1 and N2 is less typical, for example *baby–wine.* The second point of interest and the second factor to be manipulated in the experiment is the association of the verb with the N1-N2 pair. This is of interest because for an unassociated N1-N2 pair, the combination of the verb with noun pair can render the sentence as either typical or atypical. For example, the atypical combination of *baby–wine* at first glance appears to be an atypical pair. However, this unexpectedness can potentially be reversed with a verb that creates a typical event, for example *baby–wine–knock.over.* By adding the verb *knock over* to the *baby–wine* combination, the sentence potentially becomes one that is typical for readers. An appropriate verb can allow the reader to re-analyze the sentence so that it becomes one that is in line with readers' real world evens. Alternatively, the verb following an unexpected N1 and N2 pair can create a less typical event, for example *baby–wine–open.* In this case, the unexpectedness encountered at N2 may continue to carry over to the verb region.

4.1.1 Design

This study employed a 2x2 design with two factors manipulated: the typicality of noun pair and the typicality of the verb with the N1-N2 pair. Noun pair typicality refers to whether N2 was strongly or weakly associated with the subject noun. Verb typicality refers to either it is strongly or weakly associated with the noun pair. There were two levels associated with each factor with the levels fully crossed, yielding four unique conditions, as presented in Table 4.1.

		Noun pair typicality	
		Typical	Atypical
Verb typicality	Typical	**TNTV** *xiaoyinger–naiping–dafan* baby–milk.bottle–knock.over	**ANTV** *xiaoyinger–putaojiu–dafan* baby–wine–knock.over
	Atypical	**TNAV** *xiaoyinger–naiping–dakai* baby–milk.bottle–open	**ANAV** *xiaoyinger–putaojiu–dakai* baby–wine–open

TNTV = typical noun typical verb, ANTV = atypical noun typical verb, TNAV = typical noun atypical verb, ANAV = atypical noun atypical verb

Table 4.1: Factors and levels for SOV study.

Explanation of the different levels is as follows: condition TNTV is the most typical condition, where the noun pair is typical and the verb is also typical. In condition ANTV, which has an atypical noun and a typical verb, unexpectedness is encountered at N2, since it is an unexpected combination with N1. However, the verb was set up to form a

typical combination with the noun pair and there could be a reanalysis at the position of the verb. In the TNAV condition, N2 forms a typical combination with N1 but the verb is an atypical combination with the noun pair, therefore unexpectedness is not encountered until the verb. In the ANAV condition, the most atypical scenario, both N2 and the verb are atypical, therefore unexpectedness is encountered at N2 and carried over to the verb.

4.1.2 Predictions

Noun typicality

All the N1s in the *ba* sentences were animate nouns, since animate entities meet most of the features of a prototypical agent (Dowty, 1991). All the N2s in the *ba* sentences were inanimate nouns, since inanimate objects serve better as patients or the affected entity.

If an effect of noun typicality is present, it is predicted that there will be a prolonged reading time at N2 and its following regions. To assess whether there is a main effect of noun typicality, reading times at these critical regions were evaluated: N2 (the object noun), N2+1 (the region following the object noun), Verb, V+1 (the region following the verb), and V+2 (the second region after the verb). Reading times at these regions will be compared between the TNTV *(baby–milk.bottle–knock.over)* and TNAV *(baby–milk.bottle–open)* conditions versus ANTV *(baby–wine–knock.over)* and ANAV *(baby–wine–open)* conditions. Conditions TNTV and TNAV have N2s that form typical combinations with their respective N1s. These combinations are not predicted to lead to

longer processing times, as compared to conditions ANTV and ANAV where N2s are atypical and form unexpected combinations with N1. The unexpected combination, i.e., *baby–wine* will lead to a longer reading time if the typicality of a noun pair is assessed in real-time sentence processing.

Verb typicality

The typicality of the verb was manipulated to form either a typical or atypical associated with N1-N2 pair. McRae et al. (2005) showed that individual nouns generate expectations about what the subsequent verb may be. Since the verb is influenced by the preceding nouns in a SOV sentence, a verb that is not typically associated with N1 and N2 *(baby–milk.bottle–open, baby–wine–open)* is predicted to have a longer reading time than a verb that is congruent with the N1-N2 pair *(baby–milk.bottle–knock over, baby–wine–knock.over)*.

The effect of the N1-N2 association can be assessed at the post-N2 region and carried over to the verb region. In condition ANTV *(baby–wine–knock.over)*, N2 is an unexpected combination with N1, and this is where the typicality effect is expected to set in. However, the atypical reading of the noun pair can potentially be reversed by adding a verb that is associated with the N1-N2 pair, for example *baby–wine–knock.over.* On the other hand, the verb can extend the atypical interpretation, creating double violations within a sentence. For example with the combination of *baby–wine–open,* reading time is expected to slow down at the N2 position and to carry over to the subsequent regions. The critical regions to determine the effect of verb typicality are the verb and the two

regions following it. Reading times at these regions will be compared in conditions TNTV *(baby–milk.bottle–knock.over)* and ANTV *(baby–wine–knock.over)* versus conditions TNAV *(baby–milk.bottle–open)* and ANAV *(baby–wine–open)*. A longer reading time is expected at the verb and post verb regions in conditions TNAV and ANAV compared to conditions TNTV and ANTV because the verb forms an atypical event with the N1–N2 pair. The predicted onset of prolonged reading time in the four conditions is underlined in Table 4.2 below.

Condition TNTV typical noun typical verb	N1–N2–V, *baby–milk.bottle–knock.over*
Condition ANTV atypical noun typical verb	N1–<u>N2</u>–V, *baby–wine–knock.over*
Condition TNAV typical noun atypical verb	N1–N2–<u>V</u>, *baby–milk.bottle–open*
Condition ANAV atypical noun atypical verb	N1–<u>N2</u>–<u>V</u>, *baby–wine–open*

Table 4.2: Predicted longer reading time regions in SOV study.

4.1.3 Method

Materials

Based on the results of the typicality judgment studies presented in the previous chapter, specifically the SOV judgment study, 48 stimulus sets were selected from the 128 sets that were used in the judgment study. Every stimulus set contained four sentences representing the four conditions depicted in Table 4.1. A stimulus set was

selected if the scores for all sentences within the set matched closely to the overall mean scores for each condition. When selecting stimulus sets, attention was given to ensure there was considerable difference in the scores between the four conditions. For instance, condition ANAV is the most atypical scenario and tends to have the lowest score. The most typical condition TNTV needs to have the highest score amongst all conditions. Conditions with one atypical manipulation, TNAV and ANTV, were required to have a score that lie between scores of the most typical and most atypical conditions. Based on these considerations, 48 stimulus sets were chosen. Another 24 sets were also included as filler items but their scores did not have as ideal of a distribution as those chosen as target items. Stimulus sets that belong to this group often have very similar scores for conditions ANTV, TNAV, and ANAV. These were considered filler items but they have syntactic patterns that mirror those of target items. They were included because it was noted that some of these items could be considered as being semantically anomalous and the inclusion of these items adds variation to the stimulus items. The mean scores for the 48 chosen stimulus sets are presented in Table 4.3 below. The mean scores for the 24 sets of filler items that bear resemblance to the target items are presented in Table 4.4 below.

	Condition TNTV *baby–milk.bottle–knock.over*	Condition ANTV *baby–wine–knock.over*	Condition TNAV *baby–milk.bottle–open*	Condition ANAV *baby–wine–open*
Mean score (SD)	6.54 (1.00)	3.02 (1.92)	4.31 (2.16)	2.20 (1.57)

Table 4.3: Mean scores (standard deviation) of stimulus items by condition in SOV study.

	Condition TNTV *nurse–thermometer–sterilize*	Condition ANTV *nurse–soccer ball–sterilize*	Condition TNAV *nurse–thermometer–destroy*	Condition ANAV *nurse–soccer ball–destroy*
Mean score (SD)	6.46 (1.06)	2.51 (1.86)	4.67 (2.11)	2.49 (1.78)

Table 4.4 Mean scores (standard deviation) of SOV filler items by condition in SOV study.

All the target items were divided into ten regions for analysis:

(3)

N1	BA	N2	N2+1	Verb	V+1	V+2	V+3	V+4	V+5
xiaoyinger	ba	naiping	quanbu	dafan	le	yihou	jiu	pazou	le
小嬰兒	把	奶瓶	全部	打翻	了	以後	就	爬走	了。
baby	BA	milk bottle	all	knock over	PERF	then	and	crawl away	PERF

"The baby knocked over all the milk bottles and then crawled away."

All the N2+1 regions contained an adverb or a quantifier. This was necessary since the regions of N2 and V were of interest. As there could be spillover effects from critical regions (Mitchell, 1984, 2004), a buffer was needed between these two critical regions to enable the effect of noun typicality to be isolated. Without a post N2 region, the typicality of noun pair and verb would not have been able to be fully assessed, as a prolonged reading time in the verb region could be attributed to both the atypical interpretation of both the noun pair and the verb. There were five post-verb regions and they were created to allow a more natural reading of the sentences and to allow each sentence to appear as a complete sentence instead of a fragment.

All participants read 120 sentences, including 48 target items, 24 fillers that have the same structure as the targets and 48 filler items that have a variety of non-SOV syntactic structures. Two-thirds of the sentences had a comprehension question following the sentences to ensure participants were reading for comprehension. All the comprehension questions were True or False questions and were based on the content of the displayed sentence. The comprehension questions had equal numbers of "true" and "false" answers. All the sentences were distributed into four lists using a Latin Square design. Each list contained only one item from a given stimulus set and equal numbers of sentences from each condition. Participants were randomly assigned to one of the lists and items were randomized for each participant.

Procedure

This experiment was set up using the program IBEX (Drummond, 2010). IBEX is a web-based program designed for building psycholinguistic experiments whereby participants can access the experiment remotely using a web interface. Web-based experiments have gained popularity as they provide easy access to a wider target group, and studies have shown that web-based experiments and laboratory experiments produce consistent results, particularly for self-paced reading studies (Keller et al., 2009).

Participants were asked to read and give their consent for participation at the beginning of the experiment. The consent form was presented on the computer screen; subjects gave their consent by pressing a button indicating they had read and understood the procedure of the experiment. They were then asked to provide information on their

age, gender, and whether or not they were from Taiwan and learned Mandarin as their first language. As it is common for Taiwanese citizens to be bilingual in Mandarin and Taiwanese, participants were asked to provide their age at which they started learning Mandarin if they did not learn it as their first language.

The experiment began with six practice items that were similar but not identical to the experimental items. Participants then proceeded to the experiment after they had completed the practice items.

At the beginning of each trial, participants saw an asterisk '*' at the center of the screen against a white backdrop. They pressed the space bar to proceed to the next step where they saw a line of dashes on the screen, with each dash representing a Chinese character; hence the number of dashes indicated the length of the sentence. The words were displayed non-cumulatively (Just, et al. 1982). Only one word was displayed at once, and pressing the space bar reveals the next word; the previous word reverted back to dash lines.

Chinese has a logographic writing system where each written character represents a word or part of a word. In this study, whenever the participant saw a word, this could mean seeing one or more characters at the same time. For example, all nouns had two to four characters and participants saw these characters presented together on the computer screen. Characters representing the verbs were also presented together but without the perfective marker, for example, *da-fan-le* 'knocked over', the first two characters represented the verb and the third character represented the perfective marker.

Reading time was recorded at every region. Before an experimental sentence was displayed, an asterisk was displayed at the center of the screen to indicate the beginning of a new sentence. Two-thirds of the experimental sentences were followed by a comprehension question to ensure participants were reading for comprehension. They were instructed to press the "F" key if the answer was true and the "J" key if the answer was false. Participants did not receive feedback about the answer. Comprehension accuracy was used to eliminate inattentive participants and was not otherwise analyzed. The experiment session lasted about 30 minutes per participant.

Participants

This self-paced reading study was conducted at National Chengchi University in Taipei, Taiwan. Participants were recruited from two undergraduate English language classes. Participants were brought to a computer lab where they received instructions from the researcher before they proceeded with the experiment. Sixty-six (66) participants were recruited for this study (24 men, 40 women, mean age = 18.59, *SD* = 0.83). Participants received $200 NTD ($6.89 USD) for their participation.

4.1.4 Results

Data evaluation and detecting outliers

The mean comprehension accuracy on the comprehension questions was 91.15% (*SD* = 5.13). Data from two participants were excluded from analysis. One participant indicated that he/she was not from Taiwan the other participant was excluded due to low

comprehension accuracy (70%). This resulted in 64 participants whose results were included for analysis of which 16 were assigned to List 1, 19 to List 2, 17 to List 3, and 12 to List 4.

The next step in data evaluation was checking keypress time at each region for each participant over the course of the experiment to ensure that participants were performing the required task and not passively pressing keys. One of the participants' data showed that reading time leveled off for the second half of the stimuli. This was the participant who had low comprehension accuracy and whose data was removed from analysis.

All the reading times lower than 100 ms and larger than 5000 ms were discarded before any analysis was performed. The mean and standard deviation were calculated for reading time at each region. All data points that were larger than three standard deviations above the mean were replaced by a value representing three standard deviations above the mean value for the given region. This affected 2.17% of the data. The overall mean and standard deviation for the reading time in each region is presented in Table 4.5 below.

Region	N1	BA	N2	N2+1	V	V+1	V+2	V+3	V+4	V+5
Mean (SD)	441.65 (250.63)	422.39 (215.24)	475.79 (302.29)	491.39 (304.71)	520.17 (364.81)	469.75 (311.97)	398.43 (196.35)	348.79 (121.47)	386.09 (166.27)	426.52 (234.82)

Table 4.5: Overall reading time in each region for SOV study.

Method of analysis

The five critical regions of N2, N2+1, V, V+1, and V+2 were each analyzed for the effects of noun typicality and verb typicality. For each region, a linear mixed effects analysis was performed to analyze the relationship between reading time and the two manipulated factors. A mixed effects model approach for analyzing data can simultaneously control for subject and item variability without having to disregard one of the sources of variance as is the case in the traditional ANOVA subject and item analyses (Baayen et al., 2008; Winter, 2013). A full model was constructed with reading time as the response variable and noun typicality and verb typicality as fixed effects with an interaction term for these two factors. Subject and item were both random effects with random intercepts. Another model similar to the full model was constructed but without an interaction term. To determine whether the inclusion of the interaction term significantly improved the fit of the model, a likelihood ratio test of the full model with the interaction term versus the model without the interaction term was computed at each region.

For every critical region, the full model was also compared with a null model with both fixed effects removed and with a reduced model with one of the fixed effects removed. Comparing the full model with a reduced model without the fixed effect of interest shows whether the fixed effect is significant. If the comparison of the full and the reduced model's likelihoods is significant, the fixed effect is providing a significant improvement in fit.

In addition, a linear regression analysis was performed at each critical region so as to regress reading times on the offline judgment scores reported in the previous chapter. This analysis was conducted to determine if the offline scores are predictive of real-time reading. The judgment scores were standardized as z-scores. At the N2 and N2+1 regions, reading time was regressed on subject-object judgment scores. At the V, V+1, and V+2 regions, reading time was regressed on subject-object and SOV sentence judgment scores.

Noun typicality

Mean reading time and standard deviation by noun and verb typicality were determined for each region. Table 4.6 and Figure 4.1 present the mean reading time and standard deviation for each region for typical and atypical noun pairs; the difference in reading times between the two conditions is shown in Table 4.6.

	N1	BA	N2	N2+1	V	V+1	V+2	V+3	V+4	V+5
Typical nouns mean (SD)	442.98 (249.15)	428.18 (220.35)	476.08 (300.54)	481.87 (301.92)	506.48 (356.38)	460.74 (304.20)	394.06 (195.96)	347.68 (122.44)	383.00 (165.15)	423.00 (228.20)
Atypical nouns mean (SD)	440.32 (252.19)	416.58 (209.91)	475.50 (304.14)	500.93 (307.30)	533.91 (372.69)	478.78 (319.43)	402.80 (196.71)	349.91 (120.52)	389.19 (167.38)	430.10 (241.41)
Difference	2.66	11.60	0.58	-19.06	-27.43	-18.04	-8.74	-2.23	-6.19	-7.10

Table 4.6: Mean reading times by noun typicality in milliseconds.

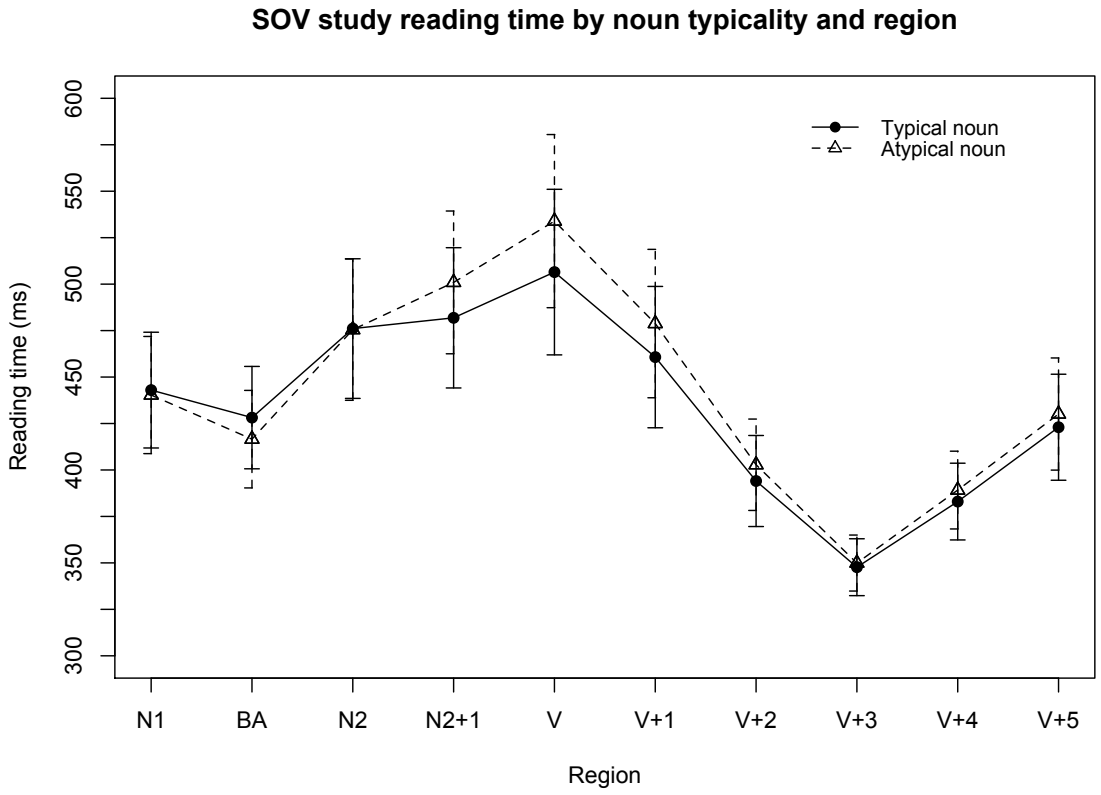

Figure 4.1: Mean reading time by noun typicality in SOV reading study. Error bars indicate one standard error below and above the mean by region.

To assess the effect of noun typicality at each of the five critical regions, a linear mixed effects analysis was conducted by comparing the full model to a reduced model with only the predictor verb typicality.

Verb typicality

The mean reading time sentences with a typical or atypical verb at each region is displayed in Table 4.7 and Figure 4.2 below.

Region	N1	BA	N2	N2+1	V	V+1	V+2	V+3	V+4	V+5
Typical items mean (SD)	452.82 (263.70)	424.88 (216.48)	476.00 (302.02)	496.92 (305.12)	523.29 (361.86)	470.03 (310.81)	397.65 (196.24)	349.80 (124.98)	383.89 (164.05)	427.84 (238.77)
Atypical items mean (SD)	429.74 (235.39)	419.72 (213.96)	475.58 (302.68)	485.50 (304.28)	516.86 (368.03)	469.45 (313.32)	399.25 (196.53)	347.72 (117.64)	388.44 (168.63)	425.08 (230.53)
Difference	23.0751	5.1544	0.4052	11.4157	6.4318	0.5810	1.6071	2.0836	-4.5511	2.7622

Table 4.7: Mean reading times for items with typical and atypical verb combinations.

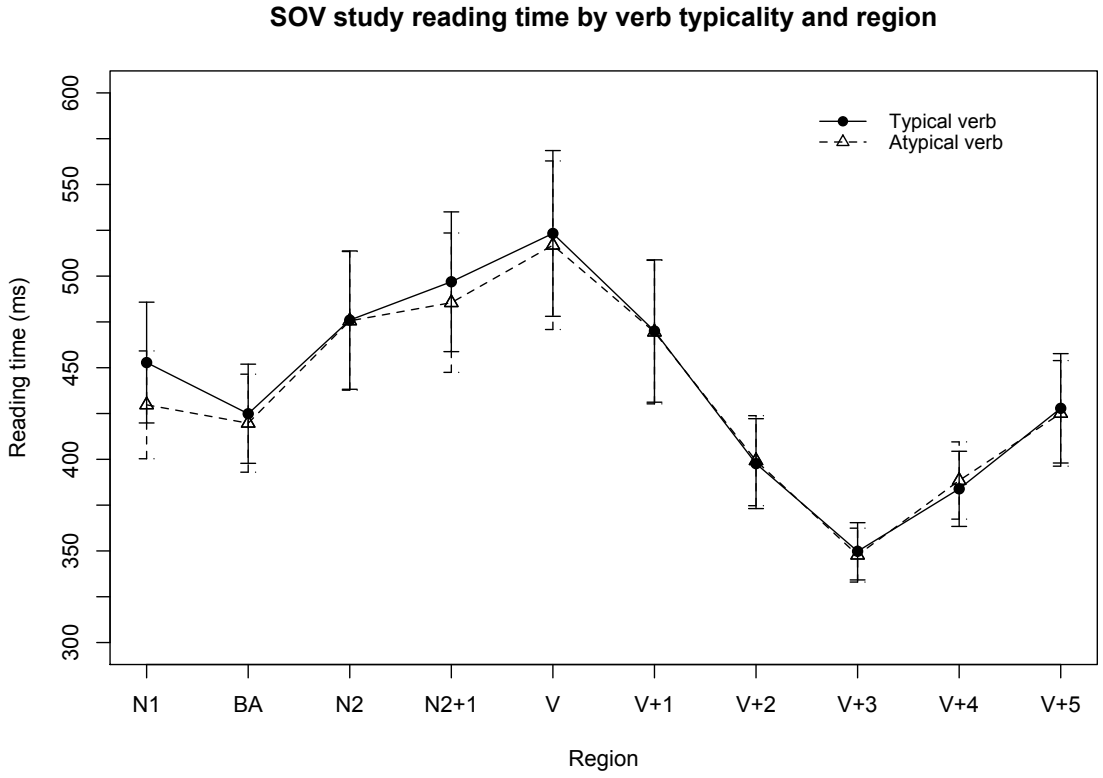

Figure 4.2: Mean reading time by verb typicality. Error bars indicate one standard error below and above the mean by region.

To determine the effect of verb typicality, a linear mixed effects analysis was performed and the full model was compared to the reduced model with only noun typicality as the predictor at the critical regions of V, V+1, and V+2.

The analysis for each critical region is reported below.

N2 Region, *naiping* "milk bottle"

A likelihood ratio test was performed to compare the full model with the interaction term versus the full model without the interaction term. Results showed that having an interaction did not significantly improve the overall fit of the model at this region, $\chi^2_{(1)} = 0.14$, $p = .71$, therefore the interaction term was not included in the full model.

There was no effect of noun typicality. Comparing the full model with a reduced model with only the predictor verb typicality did not significantly improve the fit of the full model, $\chi^2_{(1)} = .003$, $p = .95$.

A liner regression analysis was performed to regress reading time on the offline subject-object judgment scores reported in the previous chapter. A model was constructed with reading time as a function of z-transformed subject-object judgment scores. The results showed that judgment scores were not a significant predictor of reading time, $F(1, 2772) = .005$, $p = 0.94$.

N2+1 Region, *quanbu* "all"

A likelihood ratio test was performed to compare the full model with the interaction term versus the full model without the interaction term. Results showed that including the interaction term did not significantly improve the overall fit of the model for at this region, $\chi^2_{(1)} = 1.62$, $p = .20$; therefore the interaction term was not included in the full model.

At this region, inspection of the reading time differences between items with a typical noun pair and atypical noun pair suggests there could be an effect of noun typicality. Sentences with an atypical noun pair had a reading time that was on average 19.06 ms longer than sentences with a typical noun pair. A likelihood ratio test comparing the full model with a reduced model with only the predictor verb typicality approached significance, $\chi^2_{(1)} = 3.27$, $p = .07$.

A liner regression analysis was performed to regress reading time on the offline subject-object judgment scores. A model was constructed with reading time as a function of z-transformed subject-object judgment scores. The results showed that judgment scores were not a significant predictor of reading time, although it approached significance, $F(1, 2772) = 2.86$, $p = .09$. The coefficient for subject-object scores was -9.92, indicating a negative relationship between subject-object judgment scores and reading time at this region.

V Region, *dafan* "knock over"

A likelihood ratio test was performed to compare the full model with the interaction term versus the full model without the interaction term. Results showed that

having an interaction did not significantly improve the overall fit of the model for at this region, $\chi^2_{(1)} = .003$, $p = .99$; therefore the interaction term was not included in the full model.

At this region, there was an effect of noun typicality, sentences with an atypical noun pair had a reading time that was on average 27.43 ms longer than sentences with a typical noun pair. A comparison of the full model to the reduced model with only the predictor verb typicality showed that the fit of the full model was significantly greater, $\chi^2_{(1)} = 4.30$, $p < .05$.

There was no effect of verb typicality. Sentences with an atypical verb had a reading time that was on average 6 ms longer than sentences with a typical verb but this difference was not significant. Comparing the full model with a reduced model with only the predictor noun typicality did not significantly improve the fit of the full model, $\chi^2_{(1)} = .12$, $p = .73$.

A liner regression analysis was performed to regress reading time on the offline subject-object and SOV sentence judgment scores. Two separate models were constructed to regress reading time on z-transformed subject-object scores and SOV sentence scores. Creating two models were necessary as there is a significant correlation between subject-object and SOV scores, $r = .79$, $p < .001$.

A model was constructed with reading time as a function of z-transformed subject-object judgment scores. The results showed that subject-object judgment scores were not a significant predictor of reading time, although the effect of this variable approached significance, $F(1, 2772) = 3.42$, $p = .06$. A second model was constructed

with reading time as a function of z-transformed SOV judgment scores. The results showed that SOV judgment scores were a significant predictor of reading time, $F(1, 2772) = 4.01$, $p < .05$. The coefficient for SOV scores was -17.65, indicating a negative relationship between SOV judgment scores and reading time at this region.

V+1 Region, *le* Perfective marker

A likelihood ratio test was performed to compare the full model with the interaction term versus the full model without the interaction term. Results showed that the interaction term did not significantly improve the overall fit of the model at this region, $\chi^2_{(1)} = .54$, $p = .46$; therefore the interaction term was not included in the full model.

At this region, reading time differences between items with a typical noun pair and atypical noun pair suggest there could be an effect of noun typicality. Sentences with an atypical noun pair had a reading time that was on average 18.04 ms longer than sentences with a typical noun pair, and a likelihood ratio test comparing the full model with a reduced model with only the predictor verb typicality approached significance, $\chi^2_{(1)} = 2.97$, $p = .085$.

There was no effect of verb typicality. Comparing the full model with a reduced model with only the predictor noun typicality did not significantly improve the fit of the full model, $\chi^2_{(1)} = .001$, $p = .99$.

A liner regression analysis was performed to regress reading time on the offline subject-object and SOV sentence judgment scores. Two separate models were

constructed to regress reading time on subject-object scores and SOV sentence scores. A model was constructed with reading time as a function of z-transformed subject-object judgment scores. The results showed that subject-object judgment scores were not a significant predictor of reading time, $F(1, 2772) = 2.21$, $p = .14$. A second model was constructed with reading time as a function of z-transformed SOV judgment scores. The results showed that SOV judgment scores were not a significant predictor of reading time, although this variable approached significance, $F(1, 2772) = 2.79$, $p = .09$. The coefficient for SOV scores was –12.59, indicating a negative relationship between SOV judgment scores and reading time at this region.

V+2 region, *yihou* "then"

A likelihood ratio test was performed to compare the full model with the interaction term versus the full model without the interaction term. Results showed that the interaction term did not significantly improve the overall fit of the model at this region, $\chi^2_{(1)} = 1.81$, $p = .17$; therefore the interaction term was not included in the full model.

There was no effect of noun typicality. Comparing the full model with a reduced model with only the predictor verb typicality did not significantly improve the fit of the full model, $\chi^2_{(1)} = 1.57$, $p = .21$.

There was no effect of verb typicality. Comparing the full model with a reduced model with only the predictor noun typicality did not significantly improve the fit of the full model, $\chi^2_{(1)} = .10$, $p = .75$.

A liner regression analysis was performed to regress reading time on the offline subject-object and SOV sentence judgment scores. Two separate models were constructed to regress reading time on subject-object scores and SOV sentence scores.

A model was constructed with reading time as a function of z-transformed subject-object judgment scores. The results showed that subject-object judgment scores were not a significant predictor of reading time, $F(1, 2772) = 1.75$, $p = .19$. A second model was constructed with reading time as a function of z-transformed SOV judgment scores. The results showed that SOV judgment scores were not a significant predictor of reading time, although this variable approached significance, $F(1, 2772) = 3.81$, $p = .051$. The coefficient for SOV scores was -9.25, indicating a negative relationship between SOV judgment scores and reading time at this region.

Sentence-final region, *le* Perfective marker

The final position of a sentence is a region where a wrap-up effect may be evident; this is where the overall meaning and information within the sentence are fully integrated (Just et al., 1980). A linear mixed effects analysis was conducted for the sentence-final region, and the full model with the interaction term was significant compared to the full model without the interaction, $\chi^2_{(1)} = 4.04$, $p < .05$. This indicates that the inclusion of the interaction term significantly improved the fit of the model and should be included. Although there was an overall interaction effect, as illustrated in Figure 4.3, a post-hoc pairwise analysis with Tukey's method did not show any significant difference between any pairs, conditions TNTV and TNAV: $z = -1.21$, $p = .62$,

conditions TNTV and TNAV: z = -1.21, p = .62, conditions TNTV and ANTV: z = -1.99, p = .19, conditions TNTV and ANAV: z = -.29, p = .99, conditions TNAV and ANAV: z = .90, p = .81, conditions TNAV and ANAV: z = .90, p = .81, conditions ANTV and TNAV: z = .74, p = .88, and conditions ANTV and ANAV: z = 1.64, p = .36. The biggest contrast between the pairs was the difference between the most typical condition *baby–milk–knock.over* (TNTV) and the condition with an atypical noun and typical verb *baby–wine–knock.over* (ANTV), where the averaged reading times were 415.26 ms and 440.48 ms respectively. Tukey's correction method has known to be a conservative measure when there are unequal sample sizes (Whitlock & Schluter, 2009), and the contrast between conditions TNTV and ANTV was not significant with Tukey's correction, p = .19, but was significant without this correction, $p < .05$.

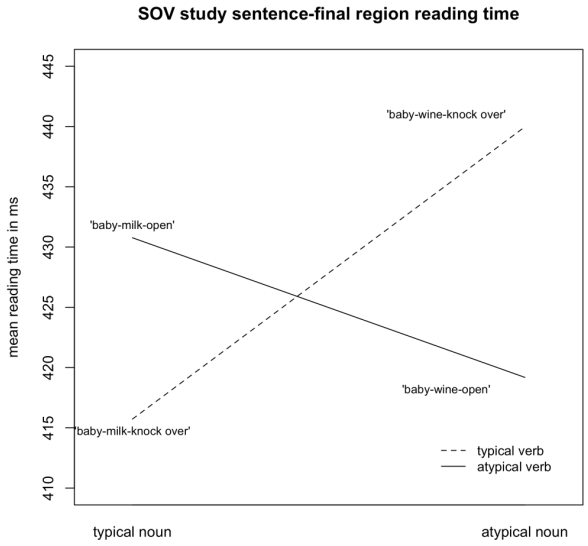

Figure 4.3: Interaction between noun and verb typicality in SOV study at sentence-final position.

Although the pairwise comparisons did not show any significant difference between the groups, this interaction effect is still worth examining since it is significant. Figure 4.3 above, Table 4.8 and Figure 4.4 below show that sentences with single violation *('baby–wine–knock.over'* and *'baby–milk–open')* had slower reading times relative to sentences expressing the most typical and least typical sentences.

Region	N1	BA	N2	N2+1	V	V+1	V+2	V+3	V+4	V+5
Condition TNTV	453.93 (259.16)	429.07 (219.57)	474.43 (302.97)	480.90 (299.27)	509.48 (354.52)	457.11 (302.73)	388.70 (189.09)	348.40 (127.35)	381.47 (164.24)	415.73 (220.30)
Condition ANTV	451.70 (268.32)	420.70 (213.42)	477.54 (301.28)	512.85 (310.22)	537.04 (368.75)	482.88 (318.34)	406.56 (202.85)	351.20 (122.65)	386.29 (163.94)	439.97 (255.53)
Condition TNAV	431.41 (237.76)	427.22 (221.32)	477.83 (298.15)	482.90 (304.91)	503.30 (358.57)	464.57 (305.93)	399.73 (202.97)	346.91 (117.10)	384.62 (166.21)	430.77 (236.28)
Condition ANAV	428.04 (233.13)	412.14 (206.13)	473.31 (307.40)	488.13 (303.85)	530.55 (377.12)	474.38 (320.77)	398.77 (189.95)	348.53 (118.27)	392.29 (171.07)	419.19 (224.47)

Table 4.8: Mean (standard deviation) reading times for conditions in SOV study.

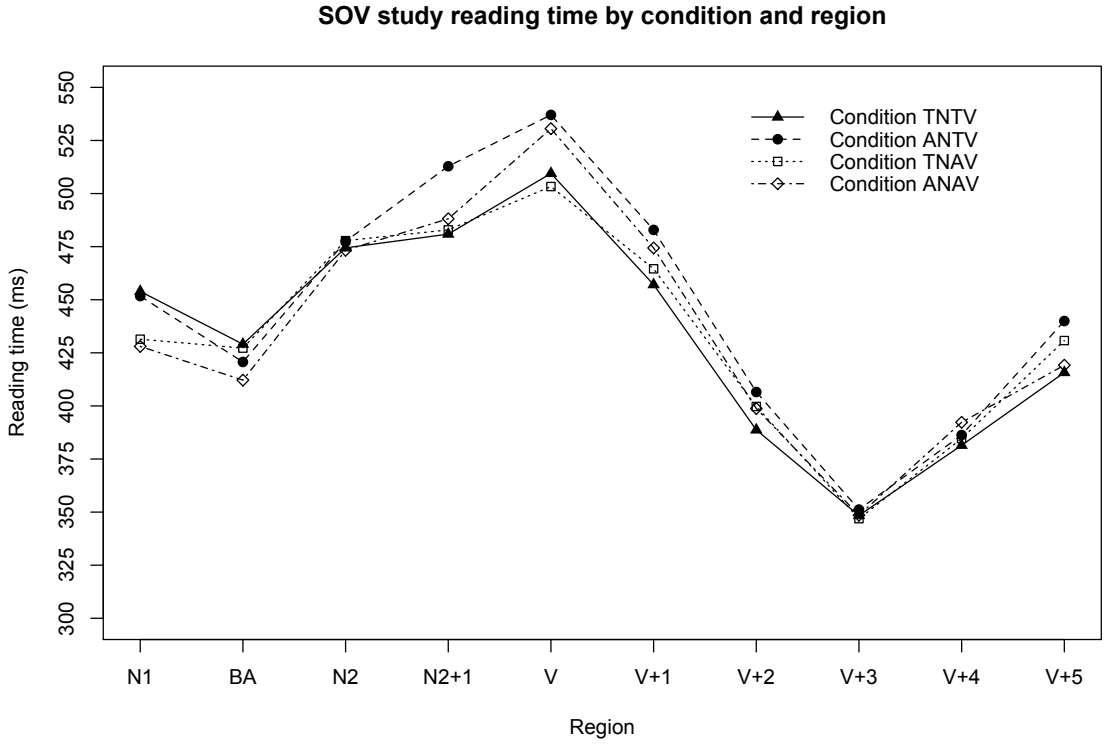

Figure 4.4: Mean reading time by condition. Error bars indicate one standard error below and above the mean by region.

The relatively faster reading time for the most typical condition *baby–milk.bottle–knock.over* was expected, as this was the condition where both the noun and verb were typical and therefore reading time should be the fastest amongst all conditions. Conditions ANTV *(baby–wine–knock.over)* and condition TNAV *(baby–milk–open)* each had one atypical item and reading times were slower than the most typical sentence and the least typical condition ANAV (*baby–wine–open*). The relatively faster reading time seen in the most atypical condition, ANAV, was not expected since atypical items usually

incur a higher processing cost which would be reflected in a slower reading time. Since the sentence-final position is not a critical region, the cause for the interaction between the two factors will not be further investigated.

4.1.5 Discussion

The results from the SOV self-paced reading study showed there was a processing cost when the nouns were weakly associated. The typicality of the verb did not contribute to additional processing cost.

The finding of a significant difference in reading time between typical and atypical nouns pairs at the verb region is most likely due to a spillover effect from N2. The effect of noun association may not show up in reading times for N2 itself; rather, it may show up in the regions following N2 as the participants attempt to integrate information from the strongly versus weakly pairs (Mitchell, 1984, 2004). In anticipation of potential spillover effects from the critical regions, the regions after the critical N2 and verb regions have, across all conditions, the same character length so as to enable comparison across conditions.

The faster reading times at the verb region when the nouns were strongly associated could be attributed to priming, especially given that the subject and object nouns appeared in a sequence without an intervening verb. Reading time was faster when the object noun was strongly associated with the subject noun, that is, when the object noun could be primed by the subject noun. However, a priming effect may not be the

only factor that contributed to this difference in reading time between strongly associated and weakly associated noun pairs. This difference in reading time, where weakly associated nouns were read slower, was not localized to the position of the object noun itself; instead, this difference did not reach significance until two regions after the object noun. Regression analysis showed that the longer reading time at the verb region was not predicted by subject-object judgment scores. When predicting reading time at the verb, subject-object scores only approached significance. Reading time at the verb region was, however, predicted by SOV judgment scores and SOV scores approached significance at the subsequent V+1 and V+2 regions, where judgment scores and reading time have a negative relationship. In summary, the faster reading time for sentences with a strongly associated noun pair cannot simply be explained by a priming effect alone; rather, the integration of the sentence's event typicality was also a contributing factor.

The results also showed that the association of the verb did not incur processing cost in real-time sentence processing. In the offline judgment study, the findings showed that people judged subject-verb pairs as more or less typically associated when they appeared in isolation. This indicates that verbs could potentially contribute incur additional processing costs. In addition, results from the regression analysis showed that the prolonged reading time at the verb region for the weakly associated noun pairs was not predicted by subject-object scores, but was predicted by SOV scores. Since the offline SOV scores were predictive of reading time, this could indicate that the prolonged reading time did not simply reflect a processing cost caused by noun pair association, but

by the typicality of the event. Verbs alone may not additional processing cost, but when they are in a phrase, they are predictive of reading time latencies.

In order to further examine the contribution of verbs in real-time sentence processing, a follow-up study was conducted using sentences with SVO word order. In the SOV study, there was no significant difference in reading time between sentences that contained a strongly or weakly associated verb. The position of the verb in the SOV study could have prevented the effect of the verb from being detected. In the SOV reading study, the verb appeared after two nouns and by the time readers reached the position of the verb, the contribution of the verb could cease to make a difference in real-time sentence processing. Given that in the offline judgment scores, people judged subject-verb pairs to be more or less typical, will this information be used in real-time sentence processing? A follow-up study with the same sentences but in SVO word order allows us to examine the contribution of verbs, since the verb in these sentences occurs immediately after the subject noun. It is possible therefore to assess the impact of the verb before the influence of the object noun. The next section reports findings from a second self-paced reading experiment to further examine the contribution of verbs in real-time sentence processing.

4.2 EXPERIMENT 2: SVO SELF-PACED READING STUDY

Results from the previous experiment showed a main effect of noun typicality but not verb typicality. The typicality of the verb did not result in differential reading times

between sentences with a typical verb and an atypical verb. The word order of the target stimuli was SOV where the two nouns preceded the verb. In order to determine the role of the verb in event typicality evaluation, a follow-up study was conducted with the experimental sentences having SVO word order. Findings from this study will show the contribution of nouns and verbs in event typicality assessment in real-time.

4.2.1 Design

This study adapted the stimuli from the SOV study with the order of the verb and N2 reversed to create sentences with SVO word order. The design of the study was also a 2x2 design with two factors manipulated: the association of N2 with N1 and the association of the verb with N1-N2 pair. There were two levels associated with each factor and the levels were fully crossed yielding four conditions, as presented in Table 4.9 below.

		Noun pair typicality	
		Typical	Atypical
Verb typicality	Typical	**Condition TVTN** *xiaoyinger–dafan–naiping* baby–knock.over–milk.bottle	**Condition TVAN** *xiaoyinger–dafan–putaojiu* baby–knock.over–wine
	Atypical	**Condition AVTN** *xiaoyinger–dakai–naiping* baby–open–milk.bottle	**Condition AVAN** *xiaoyinger–dakai–putaojiu* baby–open–wine

TVTN = typical verb typical noun, TVAN = typical verb atypical noun, AVTN = atypical verb typical noun, AVAN = atypical verb atypical noun

Table 4.9: Factors and levels for SVO self-paced reading study.

Explanation of the different levels is follows: condition TVTN was the most typical scenario, where the noun pair is typical and the verb is also typical. Condition TVTN has a typical verb and an atypical noun pair, and unexpectedness is not encountered until N2. Condition AVTN has an atypical verb and typical noun pair, and unexpectedness could be encountered at the verb. Condition AVAN was the most atypical condition, in which both the verb and N2 are atypical and unexpectedness would be encountered at the verb and carry over to the subsequent noun.

4.2.2 Predictions

Noun typicality

The main difference between the stimuli in this study and the SOV reading study was the word order. The effect of noun typicality may not be evident until the position of the N2. Although the association of the noun pair may be weakened by the intervening verb, compared to that of the SOV study, the offline judgment studies showed that subject-verb pairs were judged for their association. If the effect of noun typicality found in the SOV study holds, then it is expected that there will be a slow down in reading time for atypical N2s at the region of N2 and subsequent regions, regardless of the association of the verb. These regions will have longer reading times in conditions *baby–knock.over–wine* (TVAN) and *baby–open–wine* (AVAN), compared to reading times in conditions *baby–knock.over–milk* (TVTN) and *baby–open–milk* (AVTN).

Verb typicality

The determination of verb typicality in the SOV study was based on whether it was compatible with N1 and N2. In this study, however, the verb occurred between the two nouns. For a given subject noun, there may be few constraints on the possible set of verbs that could co-occur with it; however in the subject-verb judgment study, there was a significant difference between verbs that were considered more or less typical for a given subject. Based on the offline judgment results, the effect of verb typicality could still be evident in the reading studies. The effect of verb typicality can be assessed at the position of verb and the following regions. If there is an effect of verb typicality, we would expect there to be a prolonged reading time in the critical regions of verb and the subsequent regions in conditions *baby–open–milk.bottle* (AVTN) and *baby–open–wine*

(AVAN) relative to conditions *baby–knock.over–milk bottle* (TNTV) and *baby–knock.over–wine* (TVAN).

4.2.3 Method

Materials

The stimuli used for this study were similar to that of the SOV self-paced reading study but with a different word order. Despite the different word order, the meaning of the sentences remains the same. The construction used in the SOV construction was the Chinese *ba* construction in which the use of the *ba* marker was required; this marker is not required and not permitted in SVO sentences. The use of a pre-verbal adverb was required in the SOV sentences because there was a need for a buffer between the critical regions of N2 and the verb. This pre-verbal adverb was omitted in the SVO study since the verb immediately follows N1. A buffer between N1 and the verb was not required because N1 was not a critical region. Omitting the *ba* marker and the adverb resulted in SVO sentences having a total of eight regions, whereas the SOV sentences had ten. The eight regions of the target SVO sentences are presented in (4) below:

(4)

N1	Verb	V+1	N2	N2+1	N2+2	N2+3	N2+4
xiaoyinger	dafan	le	naiping	yihou	jiu	pazou	le
小嬰兒	打翻	了	奶瓶	以後	就	爬走	了。
Baby	knock over	PERF	milk bottle	then	and	crawl away	PERF

"The baby knocked over the milk bottle and then crawled away."

The critical regions for analysis in the SVO study were the verb (V), post-verb region (V+1), object noun (N2), post-N2 region (N2+1) and two regions following N2 (N2+2). Analyzing the V and V+1 regions will allow the effect of verb typicality to be assessed while analyzing the N2, N2+1, and N2+2 regions will allow the effect of noun typicality to be assessed.

The set up of the study was identical to that of the SOV study, where each participant read 120 sentences including 48 target items, 24 items that were fillers but have the same structure as the targets, and 48 filler items that have a variety of non-SVO syntactic structures. In line with the SOV study, two-thirds of the sentences had a comprehension question following the sentences. Participants were given six practice sentences before the experiment began. All the target sentences were distributed into four lists using a Latin Square design. Each list contained only one item from a given stimulus set. Participants were randomly assigned to a list and items were randomized for each participant.

Procedure

The procedure for this study was identical to that of the SOV self-paced reading study.

Participants

A self-paced reading study was conducted at National Chengchi University in Taipei, Taiwan. Participants were recruited from four undergraduate English language

classes. Due to time and space limitations, the experiment was conducted at a time and place of the participants' choosing. A recruiting message was distributed to students in four classes and participants accessed the experiment through a given link. A total of 79 participants were recruited for this study. Participants received $200 NTD ($6.89 USD) for their participation.

4.2.4 Results

Data evaluation and detecting outliers

The mean comprehension accuracy for the comprehension questions was 93.27% (*SD* = 4.13). Data from 74 participants were included for analyses (16 men, 58 women, mean age = 19.47, *SD* = 1.21). Three participants indicated that they were not Taiwanese and data from two participants were eliminated due to low accuracy on the comprehension questions (46.25% and 76.25%). Of the 74 participants, 16 were assigned to List 1, 17 to List 2, 22 to List 3, and 19 to List 4.

One of the participants had attempted the experiment twice. The time stamp indicated that this participant had retaken the study immediately after the first attempt. Data from the second attempt was discarded and only data from the first attempt was included for analysis.

Since this experiment was done at a time and place of the participants' own choosing, it was necessary to verify that participants were actively reading the sentences and not passively pressing the keys without reading. The next step in data evaluation was

to check for keypress time for each region in the experiment stimuli to ensure participants were reading for comprehension and not proceeding without attentive reading. After examining the data, there was no evidence showing that reading time leveled off for any of the participants.

All the reading time values less than 100 ms and greater than 5000 ms were discarded before any statistical analysis was conducted. Means and standard deviations were calculated for reading time at each region. All the data points that were over three standard deviations above the mean were replaced by a value representing 3 standard deviations plus the region's mean value. This resulted in 1.7% of the data being replaced. The overall mean and standard deviation of the reading time in each region is presented in Table 4.10 below.

Region	N1	V	V+1	N2	N2+1	N2+2	N2+3	N2+4
Mean (SD)	384.03 (204.57)	394.94 (230.92)	368.98 (179.64)	387.45 (225.74)	395.90 (233.23)	348.52 (142.72)	369.49 (182.25)	403.31 (227.38)

Table 4.10: Overall mean reading time (standard deviation) by region in the SVO study.

Method for analysis

The five critical regions of V, V+1, N2, N2+1, and N2+2 were each analyzed for the effect of noun and verb typicality. The results were analyzed with the same statistical approach as in the SOV reading study, where a linear mixed effect regression analysis was performed with fixed effects of noun and verb typicality and random intercepts for

subject and item. The procedures for determining the inclusion of the interaction term in the model, as well as assessing the significance of the fixed effects, were identical to that of the SOV reading study.

A linear regression analysis was also performed at each critical region to regress reading times on the offline judgment scores reported in the previous chapter. This analysis was conducted to examine if offline scores are predictive of real-time reading. The judgment scores were standardized into z-scores. At the V and V+1 regions, reading time was regressed on subject-verb scores. At the N2, N2+1, and N2+2 regions, reading time was regressed on subject-object, subject-verb, and SOV sentence scores.

Noun typicality

Mean reading times and standard deviations for typical and atypical noun pairs and the difference between the two conditions for every region are presented in Table 4.11 and Figure 4.5 below.

Region	N1	V	V+1	N2	N2+1	N2+2	N2+3	N2+4
Typical items mean (SD)	384.69 (202.77)	397.31 (229.77)	368.88 (176.35)	379.93 (210.84)	382.60 (213.30)	341.73 (134.72)	365.76 (180.56)	401.48 (227.59)
Atypical Items mean (SD)	383.37 (206.41)	392.56 (232.11)	369.09 (182.91)	394.96 (239.54)	409.20 (250.95)	355.30 (150.03)	373.23 (183.90)	405.15 (227.22)
Difference (SD)	1.32	4.75	-0.21	-15.03	-26.60	-13.57	-7.47	-3.67

Table 4.11: Mean (standard deviation) of items with a typical and atypical noun combination.

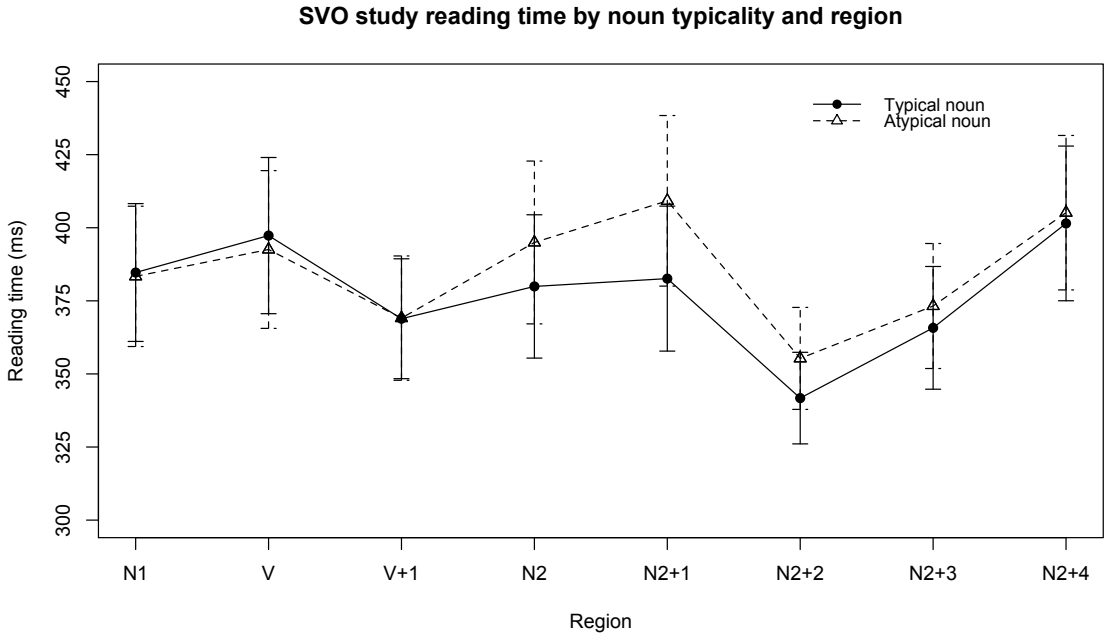

Figure 4.5: Mean reading time by noun typicality. Error bars indicate one standard error below and above the mean by region.

To analyze the effect of noun typicality, a mixed effects regression analysis was conducted by comparing the full model against a reduced model with the predictor verb typicality at each critical region.

Verb typicality

Mean reading time and standard deviation for sentences with typical and atypical verbs and the difference between the two conditions for every region are presented in Table 4.12 and Figure 4.6 below.

Region	N1	V	V+1	N2	N2+1	N2+2	N2+3	N2+4
Typical Items mean (SD)	381.99 (203.42)	391.19 (222.54)	368.46 (177.67)	389.36 (226.27)	391.45 (228.97)	347.75 (143.97)	366.93 (179.66)	403.98 (230.39)
Atypical Items mean (SD)	386.07 (205.75)	398.70 (239.04)	369.51 (181.63)	385.53 (225.26)	400.35 (237.38)	349.28 (141.50)	372.06 (184.82)	402.64 (224.41)
Difference	-4.08	-7.51	-1.05	3.83	-8.90	-1.53	-5.13	1.34

Table 4.12: Mean (standard deviation) of items with a typical and atypical verb combination.

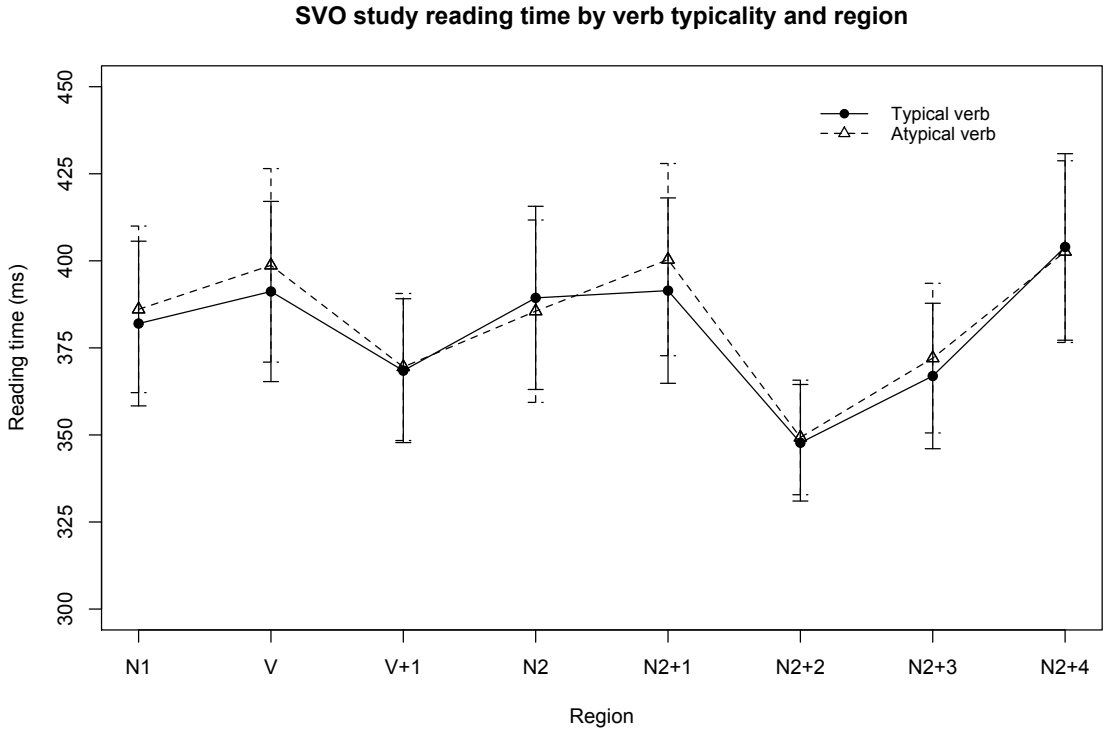

Figure 4.6: Mean reading time by verb typicality in SVO study.

Error bars indicate one standard error below and above the mean by region

To determine the effect of verb typicality, a linear mixed effects analysis was performed and the full model was compared to the reduced model with only noun typicality as the predictor at all the critical regions.

The analysis for each critical region is reported below.

V Region, *dafan* "knock over"

A likelihood ratio test was performed to compare the full model with the interaction term versus the full model without the interaction term. Results showed that having an interaction did not significantly improve the overall fit of the model for at this region, $\chi^2_{(1)} = .26$, $p = .61$, therefore the interaction term was not included in the full model.

There was no effect of verb typicality. Comparing the full model with a reduced model with only the predictor noun typicality did not significantly improve the fit of the full model, $\chi^2_{(1)} = 1.05$, $p = .31$.

A liner regression analysis was performed to regress reading time on the offline subject-verb judgment scores. A model was constructed of reading time as a function of z-transformed subject-verb judgment scores. The results showed that judgment scores were not a significant predictor of reading time, $F(1, 2840) = .15$, $p = .70$.

V+1 Region, *le* Perfective marker

A likelihood ratio test was performed to compare the full model with the interaction term versus the full model without the interaction term. Results showed that the interaction term did not significantly improve the overall fit of the model for at this region, $\chi^2_{(1)} = .18$, $p = .67$; therefore the interaction term was not included in the full model.

There was no effect of verb typicality. Comparing the full model with a reduced model with only the predictor noun typicality did not significantly improve the fit of the full model, $\chi^2_{(1)} = .02$, $p = .88$.

A liner regression analysis was performed to regress reading time on the offline subject-verb judgment scores. A model was constructed of reading time as a function of z-transformed subject-verb judgment scores. The results showed that judgment scores was not a significant predictor of reading time, $F(1, 2840) = 1.33$, $p = .25$.

N2 region, *naiping* "milk bottle"

A likelihood ratio test was performed to compare the full model with the interaction term versus the full model without the interaction term. Results showed that having an interaction did not significantly improve the overall fit of the model for at this region, $\chi^2_{(1)} = .05$, $p = .83$, therefore the interaction term was not included in the full model.

At this region, there was an effect of noun typicality, items with an atypical noun had a reading time that was on average 15.03 ms longer than items with a typical noun.

Comparing the full model with a reduced model with only the predictor verb typicality significantly improved the fit of the full model, $\chi^2_{(1)} = 4.35$, $p < .05$.

There was no effect of verb typicality. Comparing the full model with a reduced model with only the predictor noun typicality did not significantly improve the fit of the full model, $\chi^2_{(1)} = .27$, $p = .60$.

A liner regression analysis was performed to regress reading time on the offline subject-verb, subject-object, and SVO sentence judgment scores. Three separate analysis were conducted to regress reading time on subject-verb scores, subject-object scores and SVO judgment scores. These analysis were conducted separately as there was a significant correlation between subject-object and SVO scores, $r = .77$, $p < .001$, and between subject-verb and SVO scores, $r = .27$, $p < .001$.

A model was constructed with reading time as a function of z-transformed subject-verb judgment scores. The results showed that subject-verb judgment scores were not a significant predictor of reading time, $F(1, 2836) = .06$, $p = .80$. A second model was constructed with reading time as a function of z-transformed subject-object judgment scores. The results showed that subject-object judgment scores were a significant predictor of reading time, $F(1, 2836) = 4.35$, $p < .05$. The coefficient for subject-object scores was -8.07, indicating a negative relationship between subject-object judgment scores and reading time at this region. Another model was constructed with reading time as a function of z-transformed SVO judgment scores. The results showed that SVO judgment scores were a significant predictor of reading time, $F(1, 2836) = 4.01$, $p < .05$.

The coefficient for SVO scores was -9.29, indicating a negative relationship between SVO judgment scores and reading time at this region.

N2+1 Region, *yihou* "then"

A likelihood ratio test was performed to compare the full model with the interaction term versus the full model without the interaction term at the N2+1 region. Results showed that having an interaction did not significantly improve the overall fit of the model at this region, $\chi^2_{(1)}$ .29, p = .59; therefore the interaction term was not included in the full model.

There was an effect of noun typicality such that items with an atypical noun had a reading time that was on average 26.06 ms longer than items with a typical noun. Comparing the full model with a reduced model with only the predictor verb typicality significantly improved the fit of the full model, $\chi^2_{(1)} = 8.77$, $p < .01$.

There was no effect of verb typicality. Although reading time at the N2+1 region was on average 8.90 ms longer for sentences with an atypical verb, this difference was not significant. Comparing the full model with a reduced model with only the predictor noun typicality did not significantly improve the fit of the full model, $\chi^2_{(1)} = .96$, $p = .33$.

A liner regression analysis was performed to regress reading time on the offline subject-verb, subject-object, and SVO sentence judgment scores. A model was constructed with reading time as a function of z-transformed subject-verb judgment scores. The results showed that subject-verb judgment scores were not a significant predictor of reading time, $F(1, 2840) = .63$, $p = .43$. A model was constructed with

reading time as a function of z-transformed subject-object judgment scores. The results showed that subject-object judgment scores were not a significant predictor of reading time, $F(1, 2840) = 1.71$, $p = .19$. Another model was constructed with reading time as a function of z-transformed SVO judgment scores. The results showed that SVO judgment scores were a significant predictor of reading time, $F(1, 2836) = 5.16$, $p < .05$. The coefficient for SVO scores was -11.10, indicating a negative relationship between SVO judgment scores and reading time at this region.

N2+2 Region, *jiu* "and"

A likelihood ratio test was performed to compare the full model with the interaction term versus the full model without the interaction term. Results showed that the interaction term did not significantly improve the overall fit of the model for at this region, $\chi^2_{(1)} = .16$, $p = .68$; therefore the interaction term was not included in the full model.

There was an effect of noun typicality; items with an atypical noun had a reading time that was on average 13.57 ms longer than items with a typical noun. Comparing the full model with a reduced model with only the predictor verb typicality significantly improved the fit of the full model, $\chi^2_{(1)} = 11.67$, $p < .001$.

There was no effect of verb typicality. Comparing the full model with a reduced model with only the predictor noun typicality did not significantly improve the fit of the full model, $\chi^2_{(1)} = .19$, $p = .66$.

A liner regression analysis was performed to regress reading time on the offline subject-verb, subject-object, and SVO sentence judgment scores. A model was constructed with reading time as a function of z-transformed subject-verb judgment scores. The results showed that subject-verb judgment scores were not a significant predictor of reading time, $F(1, 2840) = 1.73$, $p = .19$. Another model was constructed with reading time as a function of z-transformed subject-object judgment scores. The results showed that subject-object judgment scores were not a significant predictor of reading time, $F(1, 2840) = 1.71$, $p = .19$. A third model was constructed with reading time as a function of z-transformed SVO judgment scores. The results showed that SVO judgment scores were not a significant predictor of reading time, although this variable approached significance, $F(1, 2836) = 2.80$, $p = .09$. The coefficient for SVO scores was -7.64, indicating a negative relationship between SVO judgment scores and reading time at this region.

Sentence-final region, *le* Perfective marker

At the sentence-final position, the interaction term was not significant, which was different than the result in the SOV study. As shown in Table 4.13 Figure 4.7 below, reading time at the sentence-final position was similar across all conditions.

Region	N1	V	V+1	N2	N2+1	N2+2	N2+3	N2+4
Condition TVTN	386.45 (205.43)	391.81 (218.77)	367.15 (169.38)	383.20 (212.64)	376.11 (202.91)	340.32 (131.59)	361.32 (178.26)	403.61 (237.70)
Condition TVAN	377.53 (201.41)	390.57 (226.37)	369.77 (185.68)	395.51 (239.07)	406.77 (251.51)	355.19 (155.08)	372.53 (180.97)	404.36 (222.94)
Condition AVTN	382.92 (200.17)	402.85 (240.32)	370.61 (183.13)	376.66 (209.10)	389.08 (223.09)	343.15 (137.83)	370.19 (182.81)	399.36 (217.17)
Condition AVAN	389.22 (211.25)	394.55 (237.83)	368.41 (180.21)	394.42 (240.14)	411.65 (250.51)	355.43 (144.90)	373.94 (186.90)	405.94 (231.54)

Table 4.13: Mean (standard deviation) reading times for conditions in SOV study.

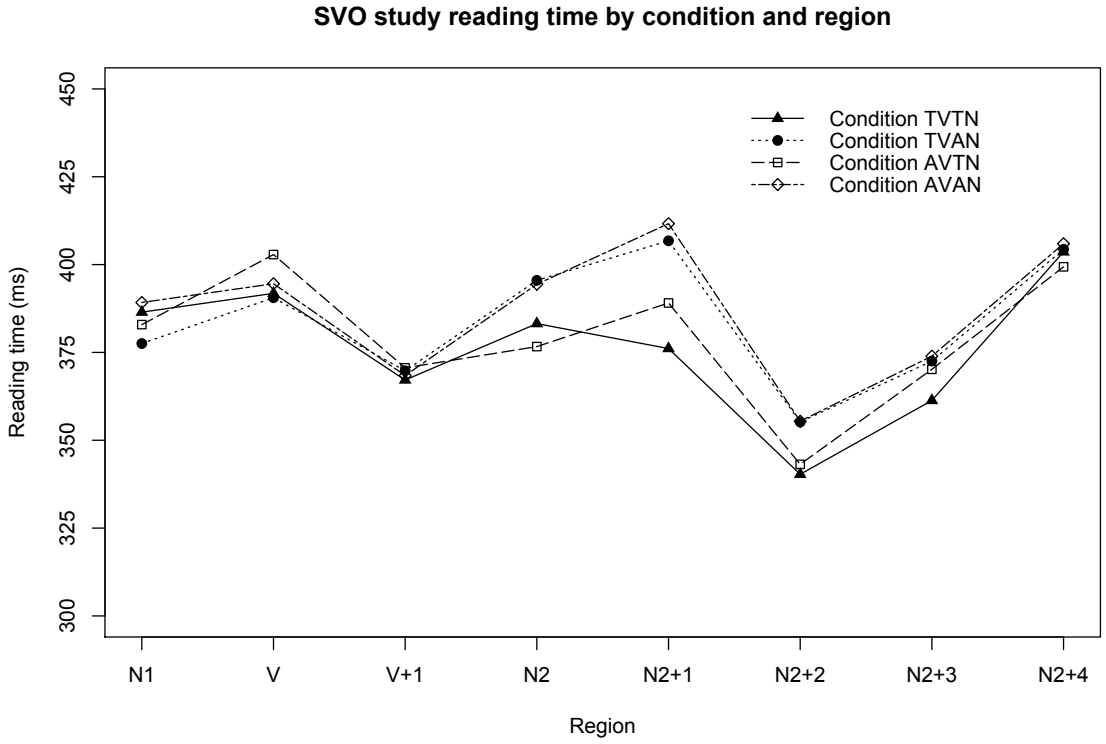

Figure 4.7: Mean reading time by condition in SVO study.

4.2.5 Discussion

The results from the SVO self-paced reading study were consistent with that of the SOV self-paced reading study. In both studies, verbs alone did not contribute to processing cost. In the offline judgment studies, people did judge subject-verb pairs as more or less typical; however this information was not used in real-time sentence processing. At the critical regions of V and V+1 in the SVO study, there was no significant difference in reading time between verbs that were strongly or weakly associated with the subject noun. In addition, regression analysis showed that subject-verb scores were not predictive of reading time at these two regions.

In the SVO study, a main effect of noun typicality was present, despite the different position of the object noun (N2) than that of the SOV study. In the SOV study, the effect of noun typicality approached significance at the N2+1 and V+1 regions, which are one region and three regions after N2, respectively. At two regions after N2, the V region, reading time for sentences with an atypical noun pair was significantly longer than for sentences with a typical one. In the SVO study, the effect of noun typicality began immediately at the object noun itself and persisted for two regions. Sentences with atypical noun pairs had reading times that were significantly longer than those with strongly associated nouns. The effect of noun association began earlier in the SVO study, compared to the SOV study, and the magnitude of the effect was also larger, as evidenced by the smaller p-values. This showed that the association of the noun pair is robust and that this association is not affected by an intervening word between the two nouns. The

faster reading time for strongly associated nouns at these three regions may suggest a priming effect. Regression analysis also showed that subject-object judgment scores were predictive of reading time at the position of the object noun. However, priming is unlikely to be the only contributor to the faster reading time, since SVO judgment scores were also predictive of reading time at the object noun position. SVO judgment scores were also predictive at the region after the object noun and approached significance at two regions after the object noun, showing that event typicality was also a contributing factor.

Comparison of SOV and SVO self-paced reading studies

Data from both the SOV and SVO studies were examined to determine whether the patterns found in the two studies were consistent. The overall mean reading times and standard deviations for the SOV study in each region are presented in the tables below.

Region	N1	BA	N2	N2+1	V	V+1	V+2	V+3	V+4	V+5
Mean (*SD*)	441.65 (250.63)	422.39 (215.24)	475.79 (302.29)	491.39 (304.71)	520.17 (364.81)	469.75 (311.97)	398.43 (196.35)	348.79 (121.47)	386.09 (166.27)	426.52 (234.82)

Table 4.5 (repeated): Overall mean reading time by region, SOV study.

Region	N1	V	V+1	N2	N2+1	N2+2	N2+3	N2+4
Mean (*SD*)	384.03 (204.57)	394.94 (230.92)	368.98 (179.64)	387.45 (225.74)	395.90 (233.23)	348.52 (142.72)	369.49 (182.25)	403.31 (227.38)

Table 4.11 (repeated) Overall mean reading time by region, SVO study.

Despite the overall faster reading time in the SVO study, the results were similar to that of the SOV study, where the main effect of noun typicality was still present and no main effect of verb typicality.

A plausible explanation for the difference in reading time can be attributed to the set up of the SVO study, where participants were given their own time and space to complete the study, whereas for the SOV study, all the participants completed the study in a computer lab. Despite the seemingly lack of control of the SVO study, cautionary measures were taken to ensure the quality of the data. To determine whether participants were reading for comprehension or simply pressing the buttons passively without reading, key press time was examined for every participant to determine whether key press time shows a reasonable fluctuation throughout the trials and over time. Results showed that there were no cases in the SVO study where participants were pressing buttons passively; therefore it can be confidently assumed that participants were indeed reading for comprehension.

Another possible explanation for the faster reading time of the SVO sentence was due to the difference in word order of the sentences in the two studies. SVO sentences are the most common type of sentences in Chinese, although the *ba* construction used for the SOV sentences is frequent in Chinese. The deviation from the canonical SVO word order of the SOV sentences is unlikely to explain the longer reading time in the SOV study. When reading time for sentence-initial noun (N1) was compared between the two studies, SVO sentences had significantly faster reading time than SOV sentences. At this position however, participants did not have cues to inform them whether this was a SVO or SOV

sentence therefore the difference in word order is unlikely to play a role in the difference in reading time. In her study, Ferreira (2003) examined how people processed sentences with non-canonical syntactic structures and found that sentences that have the agent-patient roles in a reversed order, such as passive sentences, were likely to incur a higher processing cost. For sentences that have syntactically complex structures but with canonical order, such as subject-cleft sentences, there were no additional processing costs. In the *ba* construction, the roles of agent and patient remain the same as that of a SVO sentences and therefore this construction is unlikely to incur a higher processing cost. A longer reading time for the SOV sentences was therefore unlikely to be attributed to the non-canonical syntactic structure.

4.3 SUMMARY

The chapter presented results from two online self-paced reading studies. The results in both studies were consistent, such that the association of the noun pairs had a more significant effect than the association of verbs in real-time sentence processing. The findings suggest that in real-time sentence processing, the association between the two nouns incurs a processing cost and this association is not affected by intervening words. The association of verbs by itself did not have an impact; sentences with a typical verb did not have a significantly faster reading time than sentences with an atypical verb at the critical regions. However, in both SOV and SVO studies, offline SOV and SVO scores were predictive of reading time, where less typical events were predicted to have a longer reading time. Although the association of verb by itself did not contribute to processing

cost, verbs were predictive of reading time when they occur in a phrase. The typicality of the event in a phrase, regardless of word order, was predictive of reading time.

The longer reading time at the critical regions in both SOV and SVO studies may suggest a priming effect due to the strong association between two nouns, but this may not be the only explanation. In the SOV study, a significant slow down in reading time did not occur until two regions after the object noun instead of at the object noun itself. If the slow down in reading time was solely due to the lexical association between two nouns, we would expect reading time begin to slow down at the object noun. In addition, regression analysis showed that the longer reading time was predicted by offline SOV judgment scores but not subject-object scores. In the SVO study, longer reading time began at the object noun and persisted for two regions after the object noun. In addition, regression analysis showed that at the object noun region, reading time was predicted not only by subject-object scores alone but also SVO judgment scores. Because offline SOV and SVO scores were predictive of online read times, this suggest the slow down in reading time can be attributed to the typicality of events depicted by the sentences.

The next chapter discusses the findings of the offline judgment studies and the online self-paced reading studies, and also provides possible explanations to account for the subtle influence of verbs in real-time sentence processing.

Chapter 5 Conclusion

5.1 GOALS

The goal of this dissertation was to understand whether event typicality affects sentence processing in real-time. Offline judgment studies were conducted to examine whether the association between words contributes to event typicality assessment in sentences. Online studies were conducted to determine whether judgments of word pairs and of the typicality of events described by sentences are used in real-time sentence processing. Results from these studies were analyzed to answer the following research questions:

1. Is event typicality more than the semantic association between words?
2. Does event typicality affect real-time reading?

The results from the studies reported in chapters three and four showed that the association between words does contribute to judgments of event typicality within a sentence context, and offline judgments of word association and event typicality do modulate online sentence processing. However, the association between nouns had a more significant impact on real-time sentence comprehension than the association betweens nouns and verbs did. The following sections provide a more detailed answer to the research questions.

5.2 OVERVIEW OF THE RESULTS

The results of the offline judgment studies showed that event typicality is more than the semantic association between words. For the experimental stimuli, participants judged the association between subject-object pairs and subject-verb pairs to be more or less typical. However, the difference between strongly and weakly associated noun pairs was greater than was found for subject-verb pairs, where the observed difference between the mean scores of typical and atypical subject-verb pairs was smaller than the difference between typical and atypical noun pairs.

Both the association of noun-noun and noun-verb pairs contributed to judgments of event typicality. When these word pairs were embedded in a sentence context, participants were instructed to evaluate the typicality of the event expressed by the sentence and not just word associations. Regardless of whether the word order of the sentence was SOV or SVO, the results were consistent, such that the association of the noun pairs had a more significant impact on judgments of event typicality than did verb typicality. Although there was an interaction between noun pair and verb typicality in both SOV and SVO judgment studies, the effect of verb typicality was subtler.

Judgments of word associations and event typicality modulated online sentence comprehension regardless of word order. In both SOV and SVO online self-paced reading studies, a longer reading time was observed only when the noun pair was

atypical. Offline assessments of the typicality of the events expressed by SOV and SVO clauses were also predictive of reading time.

The strong effect of noun association could be construed as meaning event typicality was less relevant in determining the processing load of sentences and that a faster reading time for strongly associated nouns was due to a priming effect. Regression analysis showed that this explanations is not adequate. In the online SOV reading study, the difference in reading times between strongly- and weakly-associated nouns did not reach significance until two regions after the object noun. Offline subject-object judgment scores were not predictive of reading time at the object noun region and the subsequent regions but SOV judgment scores were predictive at the verb region. This showed that the slower reading time was not only due to the atypical noun pair, but also to judgments of the sentence's event typicality. Similar findings were observed in the online SVO reading study. Reading times were faster at the object noun region when the noun pair was typical and continued for two more regions. It is possible that there was a priming effect, as offline subject-object judgment scores were predictive of reading time at the object noun. However, offline SVO scores were predictive of reading time at this region and also at the next region, indicating an effect of event typicality on reading times.

The studies reported in this dissertation showed that nouns had a strong impact on event typicality processing while the verbs tested in this study had a subtle effect. The next section discusses the possible explanations for the subtle effect of verbs. Certain

properties of Chinese verbs may have had an impact on the way participants processed the sentences, which will also be discussed.

5.2.1 Verbs in sentence processing

This section looks at some possible explanations for why the association between two nouns played a more significant role in assessing event typicality, and for why there was a lesser contribution from verbs. The difference between nouns and verbs needs to be considered, as they may be processed differently by the brain. Tyler et al. (2004) suggested that the reason there is a difference in the processing of nouns and verbs is not due to any difference in the representation of the noun and verb stems. Instead, it is because verbs usually have more complex inflection, resulting in stronger activity from the left inferior front gyrus, an area associated with morpho-syntactic processing. The inflections on verbs typically have a more syntactic function (e.g., tense; person & number agreement) whereas inflections for nouns tend to be more semantic (e.g., number distinction). However, this difference in English cannot explain the differences in nouns and verbs in Chinese because Chinese has simple morphology and verbs are not inflected for grammatical functions.

Verbs in Chinese have been reported to show a different pattern in Chinese children's language development than English-speaking children. Chinese children produced more verbs than nouns (Tardif, 1996); in contrast, English-speaking children produced more nouns than verbs in early stages of language development (Goldin-

Meadow et al., 1976). Ma et al. (2009) invoked the higher imageability of Chinese verbs as one of the main explanations. Imageability is the ability of a word to produce a mental image for the reader/hearer, and concrete nouns are in general more imageable than verbs and abstract nouns (Bird et al., 2003). Ma et al. argue that a Chinese verb can encode the specific manner of an action; for example, the English verb 'carry' has four counterparts in Chinese: *na* 'to carry with hands,' *bao* 'to carry with arms', *bei* 'to carry on the back,' and *duan* 'to carry flat on hand.' These verbs have narrower meanings and received higher imageability ratings. In addition, because Chinese allows pro-drop and verbs alone can be used to answer questions, verbs are more salient in Chinese than their English counterparts. However, it is not clear whether these properties of Chinese verbs have consequences for real-time sentence processing or whether there is a relationship between the typicality of the verbs in the stimuli and their imageability. Given the saliency of Chinese verbs, the next step in understanding event typicality processing could examine the processing of pro-drop sentences to determine whether verbs could make a stronger contribution in such sentences.

5.3 SUMMARY AND FUTURE WORK

Overall, the studies reported in this dissertation showed that event typicality is more than the semantic association between words, and the association between noun pairs and noun-verb pairs contribute to overall event typicality. In real-time sentence processing, the association between nouns has greater processing effects. This strong impact of noun association was not affected by an intervening word. Verbs also

contributed to event typicality processing although in a more subtle way. Offline judgments of SOV and SVO sentences, however, were predictive of reading times; longer reading times could be attributed to the atypicality of the events depicted by the stimulus sentences in question. The typicality of verbs were only predictive of reading times to the extent that verbs contributes to offline judgments of the typicality of the events expressed by SOV or SVO clauses; the judged association of subject-verb pairs was not a predictor of reading times.

The subtle contribution of the verbs alone to event typicality in real-time sentence comprehension could result from the manipulations of the verbs. In the subject-verb judgment studies, the contrast between typical and atypical pairs was smaller, compared to the contrast between typical and atypical noun pairs. The design of the offline judgment studies were centered around nouns; it is possible that a stronger effect of verbs would be detected if the studies had been designed around verbs. Further studies will be needed to examine the contribution of the verbs. In addition, since null arguments are prevalent in Chinese, the next step in understanding event typicality processing could examine the processing of pro-drop sentences to determine whether verbs could make a stronger contribution in these sentences with null arguments.

A study will be needed that seeks to determine whether this effect of noun association is sensitive to discourse context. Studies such as Nieuwland and Van Berkum. (2006) showed that unrelated words that create a meaning anomaly (e.g*., peanuts* and *fall in love*) can be accepted if they are placed in an appropriate context. In addition, a

discourse context can activate all features relevant to an event.[14] If atypical noun pairs were placed in an appropriate discourse context, these noun pairs might be interpreted as being associated. By removing the effect of noun association, the verb could be manipulated for its association with the noun pair to determine the contribution of verbs in event typicality assessment.

The effect of noun association will also need to be investigated further. In the SOV reading study, an effect of noun pair association was observed, and in a follow-up SVO reading study, this effect was still present although the nouns were separated by an intervening verb. Further studies are needed to determine whether this effect still holds when there is a greater distance between the two nouns. This will need to be tested in structurally complex sentences, such as relative clauses where filler-gap dependencies must be established among several referents. In addition, corpus studies of Chinese noun-noun and noun-verb co-occurrence frequencies will be needed to determine the relationship between word association and co-occurrence inasmuch as the strong semantic association between words may be correlated with frequency of co-occurrence (Spence & Owens, 1990). The studies reported in this dissertation used word pairs and sentences from Chinese, but the results could be compared to other languages with a dominant SOV word order and to other pro-drop languages, to further determine the role of word association in event typicality judgment.

---

[14] Metusalem et al. presented studies showing that, when a snow scene was depicted, *snowman* was the most appropriate continuation for the target sentence *the kids built a* _____. An inappropriate but relevant continuation such as *jacket* elicited a smaller N400 amplitude compared to the inappropriate and irrelevant word *towel.*

Appendices

APPENDIX A: VERB-OBJECT JUDGMENT STUDY

An offline judgment study was conducted to elicit native speaker's intuition on the association of verb-object pairs. The goal was to determine whether people judge these pairs to be different based on their association. Verb-object sequence is a possible sentence structure in Chinese as the language allows subject omissions. Given this sentence structure, participants may be judging verb-object pairs as events instead of words in isolation. The results from this study were not included for analysis since the sequence of verb-object was never presented in isolation in any of the offline or online studies.

Method

Material

The set up of this study was identical to the SVO judgment study. The materials were based on the SVO study with the subject noun omitted. There were 72 stimulus sets with four conditions in each set. For every stimulus set, a verb was either typical or atypical and each was paired with a typical object and an atypical object. The typicality of the verbs was determined by its association with the noun pairs in the SVO study. Four lists were created and the stimuli were assigned to the list using a Latin Square design.

Each list contained only one item from a given stimulus set and each list had an equal number of items from each condition.

Procedure

The processes for participant recruitment and screening were identical to the judgment studies reported in chapter 3.

Participants were asked to provide a judgment score on the verb-object pairs on a 1 to 7 scale. Participants were asked to rate on the likelihood of occurrence of the verb-noun pairs[15]. A score of 1 indicates that the verb-object pair is a highly unlikely pair, while a score of 7 indicates that the verb-object pair is a highly likely pair.

Participants

The total number of participants for this study was 51 (35 female, 16 male). List 1 had 10 participants, List 2 had 17 participants, List 3 had 11 participants, and List 4 had 13 participants. The average age of the participants was 30.67 (*SD* = 5.11, range = 20-42). One participant indicated that he or she had learned Taiwanese as their first language and started learning Mandarin at the age of six. Data from this participant was included for analysis. Participants volunteered their time without compensation.

Results

The design of the study was identical to that of the SVO study. This study also employed a three way 2x2x4 ANOVA design with two within-subject factors of noun

[15] Instructions in Chinese: 請根據您的直覺，來評估每一組動詞與名詞組合發生的可能性。

typicality and verb typicality. A between-subject factor of item list was also included to ensure there was no variation between the items amongst the lists. The mean scores and standard deviations for each condition are presented in Table A1 below.

	Condition TVTN *wash–knife*	Condition TVAN *wash–military.uniform*	Condition AVTN *mail–knife*	Condition AVAN *mail–military.uniform*
Mean (*SD*)	5.47 (1.98)	4.62 (2.29)	3.88 (2.34)	3.63 (2.30)

TVTN = typical verb typical noun, AVTN = atypical verb typical noun, TVAN = typical verb atypical noun, AVAN = atypical verb atypical noun

Table A1: Mean scores and standard deviations for each condition in the VO typicality judgment study.

ANOVA showed main effects of noun typicality, $F(1, 46) = 67.06$, $p < .001$, and verb typicality, $F(1, 46) = 326.88$, $p < .001$. The interaction between noun and verb typicality was also significant, $F(1, 46) = 19.46$, $p < .001$. Post-hoc analysis with Tukey's correction showed a significant difference between all of the four combinations of levels for noun and verb typicality, all $p < .001$. Although the mean score for conditions AVTN (*mail–knife*) and AVAN (*mail–military.uniform*) was almost identical, 3.88 and 3.63 respectively, this difference was nevertheless marginally significant, $t(46) = 2.67$, $p = .049$. While this may seem surprising, the dependent variable used rating scores with a range of 1 to 7, and the small range of the scores could mean a small difference in group means will be detected as significant.

There was no main effect of item list, $F(3, 46) = 1.00$, $p = 0.40$.There was no interaction between list and noun typicality, $F(3, 46) = 0.08$, $p = 0.97$, and list and verb typicality, $F(3, 46) = 2.12$, $p = .11$, but there was an interaction between all three predictors, $F(3, 46) = 4.80$, $p < .01$. The adjusted R-squared for this model was 0.11. Removing list variable from the model, the adjusted R-squared for the model with noun and verb typicality predictors was 0.09. The adjusted R-squared for this model is the lowest among all the adjusted R-squared of the offline judgment studies reported in chapter three. Although both noun and verb typicality were significant predictors and the scores in all four conditions were different from each other, the small adjusted R-squared indicates that the variation in judgment scores was not well explained by noun and verb typicality. Regardless of whether the verb-object pairs were judged for their typicality of association or for their event typicality, in the absence of a subject noun, participants did judge the verb-object pairs to be different. However, the contrasts in manipulated typicality were weaker than that of subject-object pairs and subject-verb pairs.

APPENDIX B: EXPERIMENTAL STIMULI FOR OFFLINE SUBJECT-OBJECT JUDGMENT STUDY

The stimuli for the offline subject-object judgment study are presented below. There were 128 sets of stimuli, with stimuli for each set presented in two rows below. The first row in each set contains a subject noun (N1), which was paired with two strongly associated object nouns (N2). The second row presents two weakly associated object nouns. For example, in set 1, the subject noun *chef* was paired with the strongly associated object nouns *knife* and *cookbook*. In contrast, the weakly associated object nouns for *chef* were *military uniform* and *grenade.* The stimulus items were presented in Chinese and always as a subject-object pair, for example 廚師--刀子 "chef-knife."

	N1		N2		N2	
1	廚師	chef	刀子	knife	食譜	cookbook
	廚師	chef	迷彩服	military uniform	手榴彈	grenade
2	油漆工	painter	梯子	ladder	滾筒	roller
	油漆工	painter	防彈衣	bullet-proof vest	警車	police car
3	小男孩	little boy	足球	soccer ball	棒球	baseball
	小男孩	little boy	報表	financial statement	稅單	tax form
4	軍人	soldier	迷彩服	military uniform	手榴彈	grenade
	軍人	soldier	稻草人	scarecrow	稻草	hay
5	解說員	narrator	展覽品	exhibition items	藝術品	art work
	解說員	narrator	溫度計	thermometer	紗布	gauze
6	收藏家	collector	名畫	famous painting	古董	antique
	收藏家	collector	合約	contract	記者會	press conference
7	演員	actor	劇本	script	電影	movie
	演員	actor	烤箱	oven	麵粉	flour
8	化妝師	make-up artist	鏡子	mirror	口紅	lipstick
	化妝師	make-up artist	指南針	compass	帳棚	tent
9	警察	police	防彈衣	bullet-proof vest	警車	police car

	警察	police	標本	specimen	培養皿	petri dish
10	郵差	postman	包裹	package	信件	letter
	郵差	postman	監視器	monitor	警棍	baton
11	服務生	waiter	餐巾	napkin	小費	tip
	服務生	waiter	攝影機	video camera	紀錄片	documentary
12	新娘	bride	嫁妝	dowry	鑽戒	diamond ring
	新娘	bride	汽油	gasoline	打火機	lighter
13	秘書	secretary	行程表	itinerary	文件	document
	秘書	secretary	試管	test tube	顯微鏡	microscope
14	公車司機	bus driver	車票	bus ticket	安全帶	seat belt
	公車司機	bus driver	刀子	knife	食譜	cookbook
15	經紀人	agent	合約	contract	記者會	press conference
	經紀人	agent	手術台	operating table	麻醉藥	anesthesia
16	小偷	thief	珠寶	jewelry	面罩	face mask
	小偷	thief	垃圾桶	trash can	灰塵	dust
17	捕手	catcher	手套	gloves	面罩	facial mask
	捕手	catcher	放大鏡	magnifying glass	指紋	finger print
18	護士	nurse	溫度計	thermometer	紗布	gauze
	護士	nurse	足球	soccer ball	棒球	baseball
19	民意代表	parliament representative	法案	legislative bill	政見	political agenda
	民意代表	parliament representative	顏料	paint	畫筆	paint brush
20	求職者	job seeker	履歷表	resume	自傳	autobiography
	求職者	job seeker	鐵絲網	barbed wire	毛毯	blanket
21	農夫	farmer	稻草人	scarecrow	稻草	hay
	農夫	farmer	簡報	presentation slides	公文	official document
22	驗光師	optometrist	眼鏡	glasses	鏡片	lens
	驗光師	optometrist	珠寶	jewelry	警報器	security alarm
23	富翁	millionaire	遊艇	boat	別墅	mansion
	富翁	millionaire	口哨	whistle	計時器	timer
24	裁縫師	seamstress	布料	fabric	扣子	button
	裁縫師	seamstress	劇本	script	電影	movie
25	畫家	painter (artist)	顏料	paint	畫筆	paint brush

	畫家	painter (artist)	電腦	computer	程式	software program
26	鎖匠	locksmith	門鎖	door lock	鑰匙	key
	鎖匠	locksmith	雜誌	magazine	報紙	newspaper
27	清潔工	cleaner	垃圾桶	trash can	灰塵	dust
	清潔工	cleaner	行程表	itinerary	文件	document
28	麵包師父	baker	烤箱	oven	麵粉	flour
	麵包師父	baker	鏡子	mirror	口紅	lipstick
29	長官	executive	簡報	presentation slides	公文	official document
	長官	executive	功課	homework	書包	schoolbag
30	難民	refugee	鐵絲網	barbed wire	毛毯	blanket
	難民	refugee	照相機	camera	紀念品	souvenir
31	飯局主人	host	葡萄酒	wine	邀請函	invitations
	飯局主人	host	車票	bus ticket	安全帶	seat belt
32	老闆	boss	計算機	calculator	商品	merchandise
	老闆	boss	化石	fossil	古蹟	historic site
33	學生	student	功課	homework	書包	schoolbag
	學生	student	名畫	famous painting	古董	antique
34	毒販	drug dealer	毒品	drug	大麻	marijuana
	毒販	drug dealer	履歷表	resume	自傳	autobiography
35	獸醫	veterinary	手術台	operating table	麻醉藥	anesthesia
	獸醫	veterinary	漁船	fisherboat	救生艇	lifeboat
36	麻醉師	anesthetist	針頭	needle	心電圖	ECG
	麻醉師	anesthetist	布料	fabric	扣子	button
37	科學家	scientist	試管	test tube	顯微鏡	microscope
	科學家	scientist	奶瓶	milk bottle	搖籃	cradle
38	牙醫	dentist	假牙	denture	口罩	mask
	牙醫	dentist	演唱會	music concert	歌曲	song
39	海巡人員	coast guard personnel	漁船	fisherboat	救生艇	lifeboat
	海巡人員	coast guard personnel	推薦信	recommendation letter	研討會	seminar
40	消防員	firefighter	滅火器	fire extinguisher	斧頭	ax
	消防員	firefighter	展覽品	exhibition items	藝術品	art work
41	教授	professor	推薦信	recommendation letter	研討會	seminar

	教授	professor	梯子	ladder	滾筒	roller
42	工程師	engineer	電腦	computer	程式	computer program
	工程師	engineer	點滴	iv bag	花束	bouquet
43	會計師	accountant	報表	financial statement	稅單	tax form
	會計師	accountant	吹風機	hair dryer	剪刀	scissors
44	記者	reporter	訪談	interview	新聞稿	press release
	記者	reporter	成本	cost	工廠	factory
45	偵探	detective	放大鏡	magnifier	指紋	fingerprint
	偵探	detective	計算機	calculator	商品	merchandise
46	縱火犯	arsonist	汽油	gasoline	打火機	lighter
	縱火犯	arsonist	積蓄	savings	存款	deposit
47	模特兒	model	高跟鞋	high-heeled shoes	項鍊	necklace
	模特兒	model	擴音器	amplifier	急救箱	first aid kit
48	探險家	explorer	指南針	compass	帳棚	tent
	探險家	explorer	高跟鞋	high-heeled shoes	項鍊	necklace
49	理髮師	barber	吹風機	hair dryer	剪刀	scissors
	理髮師	barber	眼鏡	glasses	鏡片	lens
50	病人	patient	點滴	iv bag	花束	bouquet
	病人	patient	嫁妝	dowry	鑽戒	diamond ring
51	考古學家	archaeologist	化石	fossil	古蹟	historic site
	考古學家	archaeologist	毒品	drug	大麻	marijuana
52	裁判	referee	口哨	whistle	計時器	timer
	裁判	referee	包裹	package	信件	letter
53	聲樂家	vocalist	演唱會	music concert	歌曲	song
	聲樂家	vocalist	門鎖	door lock	鑰匙	key
54	詐騙集團	con artist	積蓄	savings	存款	deposit
	詐騙集團	con artist	法案	legislative bill	政見	political agenda
55	小嬰兒	baby	奶瓶	milk bottle	搖籃	cradle
	小嬰兒	baby	葡萄酒	wine	邀請函	invitations
56	保全人員	security guard	監視器	monitor	警棍	baton
	保全人員	security guard	訪談	interview	新聞稿	press release
57	導演	director	攝影機	camera	紀錄片	documentary

	導演	director	針頭	needle	心電圖	ECG
58	企業家	entrepreneur	成本	cost	工廠	factory
	企業家	entrepreneur	餐巾	napkin	小費	tip
59	編輯	editor	雜誌	magazine	報紙	newspaper
	編輯	editor	滅火器	fire extinguisher	斧頭	ax
60	救生員	lifeguard	擴音器	amplifier	急救箱	first aid kit
	救生員	lifeguard	手套	gloves	面罩	facial mask
61	觀光客	tourist	照相機	camera	紀念品	souvenir
	觀光客	tourist	油漆	paint	瓷磚	tile
62	裝潢工人	decorating workers	油漆	paint	瓷磚	tile
	裝潢工人	decorating workers	墨鏡	sunglasses	西裝	business suit
63	保鏢	bodyguard	墨鏡	sunglasses	西裝	business suit
	保鏢	bodyguard	假牙	denture	口罩	mask
64	生物學家	biologist	標本	specimen	培養皿	petri dish
	生物學家	biologist	遊艇	boat	別墅	mansion
65	收銀員	cashier	信用卡	credit card	發票	receipt
	收銀員	cashier	黑板	blackboard	粉筆	chalk
66	街頭藝人	street performer	銅板	coins	畫像	portrait
	街頭藝人	street performer	餐車	dining cart	救生衣	life jacket
67	技術員	technician	零件	parts	護目鏡	protective goggles
	技術員	technician	傳票	summons	保釋金	bail
68	商人	businessman	訂單	order	存貨	inventory
	商人	businessman	聽診器	stethoscope	手術刀	scalpel
69	木匠	carpenter	電鋸	chainsaw	木屑	sawdust
	木匠	carpenter	魚網	fishnet	魚餌	bait
70	保姆	nanny	嬰兒車	stroller	尿布	diaper
	保姆	nanny	頭盔	helmet	水泥	cement
71	小女孩	little girl	洋娃娃	doll	裙子	skirt
	小女孩	little girl	啤酒	beer	酒杯	glass
72	觀眾	audience	門票	ticket	節目單	playbill
	觀眾	audience	電鋸	chainsaw	木屑	sawdust

73	將軍	military general	吉普車	jeep	勳章	medal
	將軍	military general	玩具	toy	棒棒糖	lollipop
74	小孩	child	玩具	toy	棒棒糖	lollipop
	小孩	child	信用卡	credit card	發票	receipt
75	營養師	nutritionist	菜單	menu	維他命	vitamin
	營養師	nutritionist	弓箭	bow and arrow	槍枝	gun
76	鋼琴家	pianist	演奏會	recital	樂譜	music score
	鋼琴家	pianist	炸彈	bomb	炸藥	explosives
77	藥師	pharmacist	止痛藥	painkiller	白袍	white robe
	藥師	pharmacist	談判	negotiation	宴會	banquet
78	醫生	doctor	聽診器	stethoscope	手術刀	scalpel
	醫生	doctor	麥克風	microphone	大字報	poster
79	傭人	servant	洗衣機	washing machine	吸塵器	vacuum cleaner
	傭人	servant	捐款	donation	財產	asset
80	立法委員	legislator	選票	votes	預算	budget
	立法委員	legislator	肥料	fertilizer	雜草	weeds
81	運動員	athlete	球鞋	sneakers	獎杯	cup
	運動員	athlete	演奏會	recital	樂譜	music score
82	大使	ambassador	談判	negotiation	宴會	banquet
	大使	ambassador	喇叭	speakers	音響	stereo
83	牛仔	cowboy	靴子	boots	繩子	rope
	牛仔	cowboy	設計圖	design	模型	model
84	書法家	calligrapher	毛筆	brush	墨水	ink
	書法家	calligrapher	草藥	herbs	針灸	acupuncture
85	魔術師	magician	撲克牌	playing cards	氣球	balloon
	魔術師	magician	洗衣機	washing machine	吸塵器	vacuum cleaner
86	主持人	host	麥克風	microphone	大字報	poster
	主持人	host	銅板	coins	畫像	portrait
87	看護	caretaker	輪椅	wheelchair	病床	hospital bed
	看護	caretaker	鋼琴	piano	歌詞	lyrics
88	作曲家	composer	鋼琴	piano	歌詞	lyrics
	作曲家	composer	零件	parts	護目鏡	protective goggles

89	伴娘	bridesmaid	婚紗	wedding	紅包	red envelope
	伴娘	bridesmaid	球鞋	sneakers	獎杯	cup
90	雇主	employer	契約	contract	薪水	salary
	雇主	employer	洋娃娃	doll	裙子	skirt
91	錄音師	sound artist	喇叭	speakers	音響	stereo
	錄音師	sound artist	殺蟲劑	insecticide	土壤	soil
92	獵人	hunter	弓箭	bow and arrow	槍枝	gun
	獵人	hunter	氧氣瓶	oxygen bottles	珊瑚礁	coral reefs
93	老師	teacher	黑板	blackboard	粉筆	chalk
	老師	teacher	直升機	helicopter	擔架	stretcher
94	調酒師	bartender	啤酒	beer	酒杯	glass
	調酒師	bartender	地圖	map	背包	backpack
95	理財專員	financial consultant	貸款	credit	利息	interest
	理財專員	financial consultant	工具箱	toolbox	水管	pipe
96	菜販	vegetable vendor	竹籃	bamboo basket	圍裙	apron
	菜販	vegetable vendor	煙囪	chimney	禮物	gift
97	導遊	tour guide	特產	local specialties	傭金	commission
	導遊	tour guide	竹籃	bamboo basket	圍裙	apron
98	聖誕老人	Santa Claus	煙囪	chimney	禮物	gift
	聖誕老人	Santa Claus	止痛藥	painkiller	白袍	white robe
99	水電工	plumber	工具箱	toolbox	水管	pipe
	水電工	plumber	門票	ticket	節目單	playbill
100	攝影師	photographer	照片	photo	鏡頭	shot
	攝影師	photographer	輪椅	wheelchair	病床	hospital bed
101	黑道	ganster	保護費	protection fee	武器	weapon
	黑道	ganster	菜單	menu	維他命	vitamin
102	選民	voter	印章	seal	身份證	id cards
	選民	voter	羊毛	wool	拐杖	crutch
103	漁夫	fisherman	魚網	fishnet	魚餌	bait
	漁夫	fisherman	吉普車	jeep	勳章	medal
104	美容師	beautician	毛巾	towel	化妝品	cosmetic
	美容師	beautician	保護費	protection fee	武器	weapon

105	空服員	flight attendant	餐車	dining cart	救生衣	life jacket
	空服員	flight attendant	論文	dissertation	獎學金	scholarship
106	作家	writer	散文	prose	小說	novel
	作家	writer	護照	passport	簽證	visa
107	間諜	spy	機密	confidential information	竊聽器	tapping machine
	間諜	spy	嬰兒車	stroller	尿布	diaper
108	牧羊人	shepherd	羊毛	wool	拐杖	crutch
	牧羊人	shepherd	蛋糕	cake	奶油	cream
109	律師	lawyer	證據	evidence	訴狀	complaint
	律師	lawyer	手電筒	flashlight	機車	scooter
110	編劇	screenwriter	音樂劇	musical	舞台劇	theater
	編劇	screenwriter	訂單	order	存貨	inventory
111	慈善家	philanthropist	捐款	donation	財產	asset
	慈善家	philanthropist	靴子	boots	繩子	rope
112	囚犯	prisoner	手銬	handcuffs	腳鏈	ankle chain
	囚犯	prisoner	毛筆	brush	墨水	ink
113	旅行家	traveler	地圖	map	背包	backpack
	旅行家	traveler	契約	contract	薪水	salary
114	建築師	architect	設計圖	design	模型	model
	建築師	architect	散文	prose	小說	novel
115	檢查官	district attorney	傳票	summons	保釋金	bail
	檢查官	district attorney	選票	votes	預算	budget
116	園丁	gardener	肥料	fertilizer	雜草	weeds
	園丁	gardener	望遠鏡	telescope	绯聞	gossip
117	巡邏員	patrol	手電筒	flashlight	機車	scooter
	巡邏員	patrol	特產	local specialties	傭金	commission
118	研究生	graduate student	論文	dissertation	獎學金	scholarship
	研究生	graduate student	機密	confidential information	竊聽器	tapping machine
119	中醫	Chinese medicine doctor	草藥	herbs	針灸	acupuncture
	中醫	Chinese medicine	印章	seal	身份證	id cards

		doctor				
120	狗仔隊	paparazzi	望遠鏡	telescope	緋聞	gossip
	狗仔隊	paparazzi	貸款	credit	利息	interest
121	建築工人	construction worker	頭盔	helmet	水泥	cement
	建築工人	construction worker	婚紗	wedding	紅包	red envelope
122	潛水員	diver	氧氣瓶	oxygen bottles	珊瑚礁	coral reefs
	潛水員	diver	音樂劇	musical	舞台劇	theater
123	點心師傅	pastry chef	蛋糕	cake	奶油	cream
	點心師傅	pastry chef	手銬	handcuffs	腳鏈	ankle chain
124	投資人	investor	股票	stock	資金	fund
	投資人	investor	照片	photo	鏡頭	shot
125	救難人員	rescue worker	直升機	helicopter	擔架	stretcher
	救難人員	rescue worker	股票	stock	資金	fund
126	移民官	immigration officer	護照	passport	簽證	visa
	移民官	immigration officer	撲克牌	playing cards	氣球	balloon
127	恐布分子	terrorist	炸彈	bomb	炸藥	explosives
	恐布分子	terrorist	毛巾	towel	化妝品	cosmetic
128	果農	fruit grower	殺蟲劑	insecticide	土壤	soil
	果農	fruit grower	證據	evidence	訴狀	complaint

APPENDIX C: EXPERIMENTAL STIMULI FOR OFFLINE SOV AND SVO JUDGMENT STUDIES

The stimuli for the offline SOV judgment study are presented below. There were 128 sets of stimuli with four conditions in each set. The object noun was manipulated and was either strongly or weakly associated with the subject noun. The two object nouns in each set are presented after the BA marker and separated by a slash line (/), with the strongly associated noun presented first. Each set contains a verb that denotes a typical or atypical action performed by the protagonist. The typical verb and atypical verb are separated by a slash line (/), with the typical verb presented first.

For example, in set 1, *chef* is the subject noun, followed by the marker BA. The two nouns used were *knife* and *military uniform* which were strongly and weakly associated respectively. The typical and atypical verb for this set were *washed* and *mailed,* respectively. All the stimuli items were presented in Chinese and presented as simple SOV sentences.

The same stimulus items were used for the offline SVO judgment study. In the SVO study, the BA marker was not required and was therefore omitted. The verb occurred after the subject noun, for example, 廚師[清洗好了/寄出去了][刀子/迷彩服] "the chef washed/mailed the knife/military uniform." The items that were used in the SVO judgment study are marked by an asterisk before the set number.

*	1	廚師 把 [刀子/迷彩服] [清洗好了/寄出去了]	The chef BA the knife/military uniform washed/mailed
*	2	油漆工 把 [滾筒/警車] [洗乾淨了/偷走了]	The painter BA the roller/police car washed/stole
	3	小男孩 把 [足球/報表] [弄溼了/製做好了]	The little boy BA the soccer ball/financial statement wetted/made
	4	軍人 把 [迷彩服/稻草人] [弄髒了/修剪好了]	The soldier BA the military uniform/scarecrow dirtied/tailored
	5	解說員 把 [展覽品/溫度計] [擺好了/推銷出去了]	The curator BA the exhibition item/thermometer placed/sold

	6	收藏家 把 [古董/合約] [鎖起來了/扔出去了]	The collector BA the antique/contract locked/threw away
*	7	演員把 [劇本/烤箱] [打開了/檢查好了]	The actor BA the script/oven opened/checked
*	8	化妝師 把 [鏡子/指南針] [擦亮了/丟掉了]	The makeup artist BA the mirror/compass wiped/threw out
*	9	警察把 [警車/培養皿] [清理好了/遺失了]	The policeman BA the police car/petri dish cleaned/lost
*	10	郵差 把 [包裹/警棍] [送出去了/收下了]	The postman BA the package/baton delivered/collected
*	11	服務生 把 [小費/紀錄片] [收下了/退回去了]	The waiter BA the tip/documentary accepted/returned
*	12	新娘 把 [嫁妝/汽油] [裝箱了/送出了]	The bride BA the dowry/gasoline packed/gave away
*	13	秘書 把 [文件/顯微鏡] [檢查好了/烘乾了]	The secretary BA the document/microscope checked/dried
*	14	公車司機 把 [車票/刀子] [檢查過了/偷走了]	The bus driver BA the ticket/knife checked/stole
	15	經紀人 把 [合約/麻醉藥] [銷燬了/抹乾淨了]	The agent BA the contract/anesthesia destroyed/wiped clean
*	16	小偷把 [警報器/垃圾桶] [拆掉了/準備好了]	The thief BA the security alarm/trashcan took apart/prepared
*	17	捕手把[手套/放大鏡][弄破了/消毒了]	The catcher BA the glove/magnifying glass broke/disinfected
*	18	護士 把 [溫度計/足球] [消毒了/破壞了]	The nurse BA the thermometer/soccer ball disinfected/destroyed
	19	民意代表 把[政見/畫筆] [拿出來了/換掉了]	The parliament representative BA the policy/paintbrush representative presented/changed
	20	求職者 把 [履歷表/毛毯] [寄出去了/弄破了]	The job seeker BA the resume/blanket mailed/destroyed
*	21	農夫 把 [稻草人/簡報] [做好了/寄出去了]	The farmer BA the scarecrow/presentation slides made/sent
	22	驗光師 把 [鏡片/警報器] [壓壞了/拆下來了]	The optometrist BA the lenses/security alarm broke/dismantled
*	23	富翁把別 [墅買/口哨] [買下來了/清洗好了]	The millionaire BA the mansion/whistle bought/cleaned
	24	裁縫師 把 [扣子/劇本] [拆下來了/壓壞了]	The seamstress BA the button/script tore down/broke
*	25	畫家 把 [畫筆/程式] [買下來了/鎖起來了]	The painter BA the paintbrush/software program bought/locked
*	26	鎖匠 把 [門鎖/雜誌] [破壞了/封起來]	The locksmith BA the door door lock/magazine destroyed/sealed up
*	27	清潔工 把 [垃圾桶/文件] [整理好了/燒掉了]	The cleaner BA the trashcan/documents tidied/burned
*	28	麵包師父 把 [烤箱/鏡子] [弄髒了/燒壞了]	The baker BA the oven/mirror dirtied/burned
	29	長官 把 [簡報/書包] [評論過了/製作好了]	The executive BA the presentation slides/schoolbag criticized/made
*	30	難民 把 [毛毯/紀念品] [收下了/保養好了]	The refugee BA the blanket/souvenir accepted/maintained

	31	飯局主人 把 [葡萄酒/車票] [拿出來了/裝箱了]	The host BA the wine/ticket took out/packed
	32	老闆 把 [商品/古蹟] [包起來了/踩壞了]	The boss BA the merchandise/historical site covered/stepped on
*	33	學生 把 [書包/古董] [弄丟了/賣掉了]	The student BA the school bag/antique lost/sold
*	34	毒販 把 [毒品/履歷表] [製作好了/沒收了]	The drug dealer BA the drug/resume made/confiscated
*	35	獸醫 把 [麻醉藥/救生艇] [準備好了/推銷出去了]	The veterinary BA the anesthesia/lifeboat prepared/sold
	36	麻醉師 把 [針頭/扣子] [燒掉了/弄髒了]	The anesthetist BA the needle/button burned/dirtied
	37	科學家 把 [顯微鏡/奶瓶] [消毒了/分解了]	The scientist BA the microscope/milk bottle disinfected/dismantled
	38	牙醫師 把 [假牙/歌曲] [製做好了/偷走了]	The dentist BA the denture/song finished making/stole
	39	海巡人員 把 [救生艇/推薦信] [拉上來了/裝飾好了]	The lifeguard BA the lifeboat/letter of recommendation pulled up/decorated
*	40	消防員 把 [滅火器/展覽品] [檢查好了/鎖起來]	The fireman BA the fire extinguisher/exhibition item inspected/locked up
*	41	教授 把 [推薦信/滾筒] [寄出去了/晾乾了]	The professor BA mailed/dried the letter of recommendation/roller.
	42	工程師 把 [程式/點滴] [移走了/佔有了]	The programmer BA the software program/IV bag removed/withheld
*	43	會計師 把 [報表/吹風機] [審查過了/丟掉了]	The accountant BA the financial statement/hairdryer audited/discarded
	44	記者 把 [新聞稿/成本] [整理好了/批准了]	The reporter BA the news report/cost prepared/approved
	45	偵探 把 [放大鏡/商品] [擦乾淨了/製作好了]	The detector BA the magnifying glass/merchandise cleaned/made
	46	縱火犯 把 [汽油/存款] [偷走了/檢查好了]	The arsonist BA stole/checked
	47	模特兒 把 [高跟鞋/急救箱] [踩壞了/裝飾完了]	The model BA the gasoline/savings broke/decorated the high heels/first aid kit
*	48	探險家 把 [指南針/高跟鞋] [遺失了/燒掉了]	The explorer BA the compass/high heels lost/burned
	49	理髮師 把 [吹風機/鏡片] [燒壞了/打蠟了]	The hairdresser BA the hairdryer/lenses burned/waxed
	50	病人 把 [點滴/嫁妝] [移動了/洗乾淨了]	The patient BA the IV bag/dowry moved/cleaned
*	51	考古學家 把 [古蹟/毒品] [挖出來了/破壞了]	The archeologist BA the historical site/drug excavated/destroyed
*	52	裁判 把 [口哨/包裹] [抹乾淨了/拍照了]	The referee BA the whistle/package cleaned/photographed
	53	聲樂家 把 [歌曲/門鎖] [淘汰了/整理好了]	The vocalist BA the song/door lock got rid of/tidied up
	54	詐騙集團 把 [存款/政見] [偷走了/弄丟了]	The con artist BA the savings/policy stole/lost
*	55	小嬰兒 把 [奶瓶/葡萄酒] [打翻了/打開了]	The baby BA the milk bottle/wine spilled/opened

*	56	保全人員 把 [警棍/新聞稿] [烘乾了/藏起來了]	The security guard BA the baton/news report dried/hid
	57	導演 把 [紀錄片/針頭] [拍攝好了/換掉了]	The director BA the documentary/needle finished filming/replaced
	58	企業家 把 [成本/小費] [增加了/忽略了]	The entrepreneur BA the cost/tip increased/ignored
	59	編輯 把 [雜誌/滅火器] [審核完了/弄碎了]	The editor BA the magazine/extinguisher checked/shattered
*	60	救生員 把 [急救箱/手套] [換掉了/拍賣了]	The lifeguard BA the first aid kit/glove changed/auctioned
*	61	觀光客 把 [紀念品/油漆] [挑選好了/丟掉了]	The tourist BA the souvenir/paint selected/threw away
	62	裝潢工人 把 [油漆/西裝] [準備好了/埋起來了]	The decorating workers BA the paint/suit prepared/buried
*	63	保鏢 把 [西裝/假牙] [套上了/做好了]	The bodyguard BA the suit/denture put on/made
*	64	生物學家 把 [培養皿/別墅] [消毒了/破壞了]	The biologist BA the petri dish/mansion disinfected/destroyed
*	65	收銀員 把 [發票/粉筆] [丟掉了/晾乾了]	The cashier BA the receipt/chalk threw away/dried
*	66	街頭藝人 把 [銅板/救生衣] [撿起來了/丟出去了]	The street performer BA the coins/life vest picked up/tossed out
*	67	技術員 把 [零件/傳票] [檢查過了/寄出去了]	The technician BA the parts/summons inspected/mailed
*	68	商人 把 [訂單/聽診器] [檢查好了/燒掉了]	The businessman BA the order sheet/stethoscope checked/burned
*	69	木匠 把 [木屑/魚網] [烘乾了/挑選好了]	The carpenter BA the sawdust/fishnet dried/selected
	70	保母 把 [尿布/水泥] [扔掉了/拍攝好了]	The nanny BA the diaper/cement discarded/photographed
*	71	小女孩 把 [洋娃娃/酒杯] [藏起來了/烘乾了]	The little girl BA the doll/wine glass hid/dried
*	72	觀眾 把 [門票/木屑] [收起來了/撕掉了]	The audience BA the ticket/sawdust put away/tore up
*	73	將軍 把 [勳章/棒棒糖] [鎖起來了/拍賣了]	The military general BA the medal/lollipop locked/auctioned
*	74	小孩 把 [棒棒糖/發票] [吞下去了/製作好了]	The child BA the lollipop/receipt swallowed/made
*	75	營養師 把 [維他命/槍枝] [賣掉了/拍攝好了]	The nutritionist BA the vitamins/gun sold/photographed
*	76	鋼琴家 把 [樂譜/炸彈] [放好了/挖出來了]	The pianist BA the music score/bomb placed/dug up
	77	藥師把[白袍/談判][破壞了/批准了]	The pharmacist BA the lab coat/negotiation ruined/approved
*	78	醫生 把 [聽診器/麥克風] [消毒了/鎖起來了]	The doctor BA the stethoscope/microphone disinfected/locked away
	79	傭人 把 [洗衣機/捐款] [整理好了/淘汰了]	The maid BA the washing machine/donations tidied up/got rid of
	80	立法委員 把 [預算/肥料] [退回去了/增加了]	The legislator BA the budget/fertilizer

returned/increased

* 81 運動員把[獎杯/樂譜][裝箱了/丟掉了] The athlete BA the trophy/music score packed/threw away

82 大使 把 [談判/音響] [取消了/弄亂了] The ambassador BA the negotiation/stereo cancelled/messed up

* 83 牛仔 把 [靴子/設計圖] [烘乾了/剪掉了] The cowboy BA the boots/design paper dried/cut up

* 84 書法家 把 [毛筆/草藥] [晾乾了/消毒了] The calligrapher BA the paintbrush/herbs dried/disinfected

85 魔術師 把 [撲克牌/洗衣機] [藏起來了/擦亮了] The magician BA the playing cards/washing machine hid/wiped

* 86 主持人 把 [麥克風/銅板] [弄髒了/消毒過了] The host BA the microphone/coins dirtied/disinfected

* 87 看護 把 [輪椅/鋼琴] [擦乾淨了/買下來了] The caretaker BA the wheelchair/piano cleaned/bought

* 88 作曲家 把 [鋼琴/零件] [調整好了/拆掉了] The composer BA the piano/parts adjusted/dismantled

* 89 伴娘 把 [婚紗/獎杯] [收起來了/丟掉了] The bridesmaid BA the wedding gown/trophy put away/threw away

90 雇主 把 [契約/洋娃娃] [寄出去了/找出來了] The employer BA the contract/doll mailed/found

91 錄音師 把 [音響/殺蟲劑] [打開了/解開了] The sound engineer BA the stereo/insecticide opened/dismantled

* 92 獵人 把 [槍枝/氧氣瓶] [測試過了/沒收了] The hunter BA the gun/oxygen bottle tested /confiscated

* 93 老師 把 [粉筆/擔架] [弄斷了/賣掉了] The teacher BA the chalk/stretcher broke apart/sold

* 94 調酒師 把 [酒杯/背包] [淘汰了/審核完了] The bartender BA the wine glass/backpack got rid of/approved

95 理財專員 把 [貸款/水管] [推銷出去了/檢查過了] The financial consultant BA the mortgage/hose sold/inspected

96 菜販 把 [竹籃/煙囪] [丟棄了] The vendor BA the bamboo basket/chimney threw away/dismantled

* 97 導遊 把 [特產/竹籃] [推銷出去了/藏起來了] The tour guide BA the local specialties/basket sold/hid

98 聖誕老人 把 [煙囪/白袍] [弄壞了/賣掉了] Santa Claus BA the chimney/lab coat dirtied/sold

* 99 水電工 把 [水管/門票] [剪壞了/融化了] The plumber BA the hose/ticket cut/melted

100 攝影師 把 [鏡頭/輪椅] [拆掉了/消毒了] The photographer BA the camera lens/wheelchair took down/disinfected

101 黑道 把 [保護費/維他命] [收齊了/燒掉了] The gangster BA the protection fee/vitamins collected/burned

* 102 選民 把 [印章/羊毛] [找出來了/扔掉了] The voter BA the personal seal/wool looked for /threw away

* 103 漁夫 把 [魚網/勳章] [扔出去了/割壞了] The fisherman BA the fishnet/medal threw out/cut and broke

104 美容師 把 [化妝品/保護費] [疊起來了/燒掉了] The beautician BA the makeup/protection fee stacked/burned

105 空服員 把 [救生衣/論文] [拿出來了/包起來了] The flight attendant BA the life vest/dissertation took out/wrapped up

* 106 作家 把 [小說/簽證] [核對完了/燒掉了] The writer BA the novel/visa

				audited/burned
	107	間諜	把 [竊聽器/尿布] [檢查過了/清理好了]	The spy BA the tapping machine/diaper inspected/cleaned
	108	牧羊人	把 [羊毛/蛋糕] [修剪好了/上色了]	The shepherd BA the wool/cake trimmed/colored
	109	律師	把 [證據/手電筒] [分析好了/銷燬了]	The lawyer BA the evidence/flashlight analyzed/destroyed
*	110	編劇	把 [舞台劇/訂單] [完成了/拍攝好了]	The screenwriter BA the musical/order sheet finished/photographed
	111	慈善家	把 [捐款/靴子] [清算了/收齊了]	The philanthropist BA the donation/boots checked/collected
	112	囚犯	把 [手銬/毛筆] [解開了/扔出去了]	The prisoner BA the handcuff/paintbrush opened/threw out
*	113	旅行家	把 [背包/契約] [檢查過了/丟棄了]	The traveler BA the backpack/contract checked/discarded
*	114	建築師	把 [設計圖/小說] [撕掉了/剪掉了]	The architect BA the design paper/novel tore/cut up
	115	檢察官	把 [傳票/預算] [簽名了/收下了]	The prosecutor BA the summons/budget signed/accepted
	116	園丁	把 [肥料/誹聞] [準備好了/分析了]	The gardener BA the fertilizer/rumor prepared/analyzed
	117	巡邏員	把 [手電筒/特產] [打開了/破壞了]	The patrol guard BA the flashlight/local specialties opened/destroyed
*	118	研究生	把 [論文/竊聽器] [交出去了/烘乾了]	The graduate student BA the dissertation/tapping machine turned in/dried
	119	中醫師	把 [草藥/印章] [洗乾淨了/製做好了]	The traditional Chinese medicine doctor BA the herbs/personal seal washed/made
	120	狗仔隊	把 [誹聞/貸款] [挖出來了/檢查過了]	The paparazzi BA the rumor/mortgage dug up/verified
*	121	建築工人	把 [水泥/婚紗] [晾乾了/染色了]	The construction worker BA the cement/wedding gown dried/dyed
	122	潛水員	把 [氧氣瓶/舞台劇] [檢查過了/評論過了]	The diver BA the oxygen bottle/musical inspected/criticized
*	123	點心師父	把 [蛋糕/手銬] [賣掉了/買下來了]	The pastry chef BA cake/handcuff sold/bought the
*	124	投資人	把 [股票/鏡頭] [賣掉了/交出去了]	The investor BA the stock/camera sold/handed in
*	125	救難人員	把 [擔架/股票] [準備好了/遺失了]	The rescue worker BA the stretcher/stock prepared/lost
*	126	移民官	把 [簽證/撲克牌] [沒收了/撕掉了]	The immigration officer BA the visa/playing cards confiscated/tore up
*	127	恐怖份子	把 [炸彈/化妝品] [製做好了/洗乾淨了]	The terrorist BA the bomb/makeup made/washed
	128	果農	把 [殺蟲劑/證據] [拿出來了/弄髒了]	The fruit grower BA the insecticide/evidence prepared/dirtied

APPENDIX D: EXPERIMENTAL STIMULI FOR OFFLINE SV JUDGMENT STUDY

The stimuli for the offline subject-verb judgment study are presented below. There were 72 sets of stimuli and each set contains a subject noun and paired with two verbs. The verbs were either a typical action for the protagonist or an atypical action. The verb denoting a typical action is listed first. For example, in set 1, the subject noun *chef* was paired with the typical verb *wash* and atypical verb *mail.* The stimulus items were presented in Chinese and always as a subject-object pair, for example 廚師--清洗 "chef-wash."

1	廚師--清洗	chef–wash	廚師--寄出	chef – mail
2	演員--打開	actor – open	演員--檢查	actor – check
3	郵差--送出	postman – deliver	郵差--收下	postman – collect
4	警察--清理	policeman – clean	警察--遺失	policeman – (to) loose (something)
5	服務生--收下	waiter – accept	服務生--退回	waiter – return
6	公車司機--檢查	bus driver – check	公車司機--偷走	bus driver – steal
7	秘書--檢查	secretary – check	秘書--烘乾	secretary – dry
8	捕手--弄破	catcher – break	捕手--消毒	catcher – disinfect
9	農夫--做好	farmer – make	農夫--寄出	farmer – send
10	富翁--買下	millionaire – buy	富翁--清洗	millionaire – clean
11	畫家--買下	painter – buy	畫家--鎖起來	painter – (to) lock (something away)
12	鎖匠--破壞	locksmith – destroy	鎖匠--封起來	locksmith – (to) seal
13	難民--收下	refugee – accept	難民--保養	refugee – (to) maintain
14	學生--弄丟	student – (to) loose (something)	學生--賣掉	student – sell
15	毒販--製作	drug dealer – make	毒販--沒收	drug dealer – confiscate
16	獸醫--準備	veterinary – prepare	獸醫--推銷	veterinary – sell
17	消防員--檢查	fireman – inspect	消防員--鎖起來	fireman – (to) lock (something away)
18	教授--寄出	professor – mail	教授--晾乾	professor – (to) dry (something)
19	保鏢--套上	bodyguard – (to) put (something on)	保鏢--做好	bodyguard – (to) make (something)
20	探險家--遺失	explorer – (to) loose (something)	探險家--燒掉	explorer – burn

21	考古學家--挖出	archeologist – excavate	考古學家--破壞	archeologist – destroy
22	裁判--擦乾淨	referee – clean	裁判--拍照	referee – (to) photograph
23	小嬰兒--打翻	baby – spill	小嬰兒--打開	baby – open
24	觀光客--挑選	tourist – select	觀光客--丟掉	tourist – throw away
25	救生員--換掉	lifeguard – (to) change (something)	救生員--拍賣	lifeguard – (to) auction (something)
26	生物學家--消毒	biologist – disinfect	生物學家--破壞	biologist – destroy
27	小孩--吞下	child – swallow	小孩--製作	child – make
28	街頭藝人--撿起來	street performer – pick up	街頭藝人--丟出去	street performer – toss out
29	技術員--檢查	technician – inspect	技術員--寄出去	technician – mail
30	商人--檢查	businessman – check	商人--燒掉	businessman – burn
31	收銀員--丟掉	cashier – throw away	收銀員--晾乾	cashier – (to) dry (something)
32	鋼琴家--放好	pianist – (to) place (something)	鋼琴家--挖出來	pianist – dig up
33	醫生--消毒	doctor – disinfect	醫生--鎖起來	doctor – lock away
34	運動員--裝箱	athlete – pack	運動員--丟掉	athlete – (to) get rid (of something)
35	牛仔--烘乾	cowboy – (to) dry (something)	牛仔--剪掉	cowboy – (to) cut (up something)
36	書法家--晾乾	calligrapher – (to) dry (something)	書法家--消毒	calligrapher – disinfect
37	作曲家--調整	composer – adjust	作曲家--拆掉	composer – dismantle
38	伴娘--收起來	bridesmaid – put away	伴娘--丟掉	bridesmaid – throw away
39	老師--弄斷	teacher – break apart	老師--賣掉	teacher – sell
40	獵人--測試	hunter – (to) test (something)	獵人--沒收	hunter – confiscate
41	調酒師--淘汰	bartender – get rid of	調酒師--審核	bartender – (to) approve (of something)
42	導遊--推銷	tour guide – sell	導遊--藏起來	tour guide – hide
43	漁夫--扔出	fisherman – throw out	漁夫--割壞	fisherman – (to) cut and break (something)
44	作家--核對	writer – audit	作家--燒掉	writer – burn
45	看護--擦乾淨	caretaker – (to) wipe something clean	看護--買下	caretaker – buy
46	建築工人--晾乾	construction worker – (to make something) dry	建築工人--染色	construction worker – dye
47	投資人--賣掉	investor – sell	投資人--交出去	investor – (to) hand in (something)
48	移民官--沒收	immigration officer – confiscate	移民官--撕掉	immigration officer – (to) tear up (something)
49	油漆工--洗乾淨	painter – wash	油漆工--偷走	painter – steal

50	化妝師--擦亮	makeup artist – (to) wipe and shine	化妝師--丟掉	makeup artist – throw away
51	新娘--裝箱	bride – pack	新娘--送出去	bride – give away
52	小偷--拆掉	thief – take apart	小偷--準備好	thief – prepare
53	護士--消毒	nurse – disinfect	護士--破壞	nurse – destroy
54	麵包師父--弄髒	baker – (to) dirty (something)	麵包師父--燒壞	baker – burn
55	清潔工--整理	cleaner – tidy up	清潔工--燒掉	cleaner – burn
56	會計師--審查	accountant – audit	會計師--丟掉	account – discard
57	保全人員--弄髒	security guard – (to) dirty (something)	保全人員--藏起來	security guard – hide
58	木匠--烘乾	carpenter – (to) dry (something)	木匠--挑選	carpenter – select
59	小女孩--藏起來	little girl – hide	小女孩--烘乾	little girl – (to) dry (something)
60	觀眾--收起來	audience – put away	觀眾--撕掉	audience – (to) tear up (something)
61	將軍--鎖起來	military general – (to) lock (something) up	將軍--拍賣	military general – auction
62	營養師--賣掉	nutritionist – sell	營養師--拍攝	nutritionist – (to) photograph
63	主持人--弄髒	host – (to) dirty (something)	主持人--消毒	host – disinfect
64	水電工--剪掉	plumber – (to) cut (something) up	水電工--融化	plumber – melt
65	選民--找出來	voter – (to) look for (something)	選民--扔掉	voter – throw away
66	編劇--完成	screenwriter – (to) finish (something)	編劇--拍攝	screenwriter – (to) photograph
67	建築師--撕掉	architect – (to) tear up (something)	建築師--剪掉	architect – (to) cut up (something)
68	旅行家--檢查	traveler – check	旅行家--丟棄	traveler – discard
69	研究生--交出去	graduate student – turn in	研究生--烘乾	graduate student – (to) dry (something)
70	點心師父--賣掉	pastry chef – sell	點心師父--買下	pastry chef – buy
71	救難人員--準備	rescue worker – prepare	救難人員--遺失	rescue worker – (to) loose (something)
72	恐怖份子--製作	terrorist – make	恐怖份子--洗乾淨	terrorist – (to) wash and clean (something)

APPENDIX E: EXPERIMENTAL STIMULI FOR ONLINE SOV SELF-PACED READING STUDY

The stimuli for the online SOV self-paced reading study are presented below. There were 48 sets of stimuli with four conditions in each set. The object noun was manipulated and was either strongly or weakly associated with the subject noun. The two object nouns in each set are presented after the BA marker and separated by a slash line (/), where the strongly associated noun is presented first. Each set contains a verb that denotes a typical or atypical action performed by the protagonist. The typical verb and atypical verb are separated by a slash line (/), with the typical verb presented first.

For example, in set 1, *chef* is the subject noun. The two object nouns used were *knife* and *military uniform* which were strongly and weakly associated respectively. The typical and atypical verb for this set were *washed* and *mailed,* respectively. All the stimuli items were presented in Chinese and presented as SOV sentences and embedded into a longer sentence to allow a more natural reading.

1	廚師 把 [刀子/迷彩服] 迅速 [清洗好/寄出去] 了以後，就下班了。	The chef hurriedly BA the knife/military uniform washed/mailed and then got off work.
2	演員 把 [劇本/烤箱] 小心的 [打開/檢查好] 了以後，就離開了。	The actor BA carefully the script/oven opened/checked and then left.
3	郵差 把 [包裹/警棍] 匆忙的 [送出去/收下] 了以後，就離開了。	The postman BA hurriedly the package/baton delivered/collected and then left.
4	警察 把 [警車/培養皿] 慢慢的 [清理好/遺失] 了以後，就出去了。	The policeman BA eventually the police car/petri dish cleaned/lost and then went outside.
5	服務生 把 [小費/紀錄片] 草草的 [收下/退回去] 了以後，就跑掉了。	The waiter BA hurriedly the tip/documentary accepted/returned and then ran away.
6	公車司機 把 [車票/刀子] 用心的 [檢查過/偷走] 了以後，就心安了。	The bus driver BA carefully the ticket/the knife checked/stole and then felt assured.
7	秘書 把 [文件/顯微鏡] 技巧的 [檢查好/烘乾] 了以後，就放心了。	The secretary BA skillfully the document/microscope checked/blow-dried and then felt assured.
8	捕手 把 [手套/放大鏡] 無意的 [弄破/消毒好] 了之後，就去休息了。	The catcher BA carelessly the glove/magnifying glass broke/disinfected and then went to take a rest.

9	農夫 把 [稻草人/簡報] 恭敬的 [做好/寄出去] 了以後，就收工了。	The farmer BA deferentially the scarecrow/presenation slides prepared/sent and then packed up.
10	富翁 把 [別墅/口哨] 勉強的 [買下來/清洗好] 了之後，就很後悔。	The millionaire BA hesitantly the mansion/whistle bought/cleaned and then regretted it.
11	畫家 把 [畫筆/程式] 無意的 [買下來/鎖起來] 了之後，就開工了。	The painter BA mindlessly the paintbrush/software program bought/locked away and then started to work.
12	鎖匠 把 [門鎖/雜誌] 公開的 [破壞/封起來] 了之後，就跑走了。	The locksmith BA openly the door lock/magazine destroyed/sealed and then ran away.
13	難民 把 [毛毯/紀念品] 感激的 [收下/保養好] 了之後，就走掉了。	The refugee BA gratefully the blanket/souvenir accepted/maintained and then walked away.
14	學生 把 [書包/古董] 粗心 [弄丟/賣掉] 了以後，就很憂心。	The student BA mindlessly the school bag/antique lost/sold and then felt worried.
15	毒販 把 [毒品/履歷表] 偷偷的 [製作好/沒收] 了以後，就躲起來了。	The drug dealer BA secretly the drug/resume made/confiscated and then went to hide.
16	獸醫 把 [麻醉藥/救生艇] 順利的 [準備好/推銷出去] 了以後，就去休息了。	The vet BA the anesthesia/lifeboat readily prepared/sold and then went to rest.
17	消防員 把 [滅火器/展覽品] 一個不漏的 [檢查好/鎖起來] 了以後，就放心了。	The fireman BA all the fire extinguisher/exhibition item inspected/locked away and then felt assured.
18	教授 把 [推薦信/滾筒] 無理的 [寄出去/晾乾] 了以後，就去上課了。	The professor BA mindlessly the recommendation letter/roller mailed/dried and then went to class.
19	保鏢 把 [西裝/假牙] 公然的 [套上/做好] 了以後，就去上班了。	The bodyguard BA openly the suit/denture put on/made and then went to work.
20	探險家 把 [指南針/高跟鞋] 不小心 [遺失/燒掉] 了以後，就很困擾。	The explorer BA accidentally the compass/heels lost/burned and felt worried.
21	考古學家 把 [古蹟/毒品] 用力的 [挖出來/破壞] 了以後，就離開了。	The archeologist BA laboriously the historical site/drug excavated/destroyed and then left.
22	裁判 把 [口哨/包裹] 隨意的 [擦乾淨/拍照] 了以後，就去工作了。	The referee BA carelessly the whistle/package cleaned/photographed and then went to work.
23	小嬰兒 把 [奶瓶/葡萄酒] 全部 [打翻/打開] 了以後，就爬走了。	The baby BA the entire milk bottle/wine spilled/opened and then crawled away.
24	觀光客 把 [紀念品/油漆] 馬上 [挑選好/丟掉] 了以後，就去拍照了。	The tourist BA hurriedly the souvenir/paint selected/threw away and then went to take a picture.
25	救生員 把 [急救箱/手套] 安心的 [換掉/拍賣] 了以後，就去巡察了。	The lifeguard BA assuredly the first aid kit/glove changed/auctioned and then went to patrol.

26	生物學家 把 [培養皿/別墅] 從頭到尾 [消毒過/破壞] 了以後，就休假去了。	The biologist BA thoroughly the petri dish/mansion disinfected/destroyed and then went on vacation.
27	小孩 把 [棒棒糖/發票] 隨意的 [吞下去/製作好] 了以後，就跑走了。	The child BA carelessly the lollipop/receipt swallowed/made and then ran away.
28	街頭藝人 把 [銅板/救生衣] 畢恭畢敬的 [撿起來/丟出去]了以後，就收工了。	The street performer BA deferentially the coins/life vest picked up/tossed out and then packed up.
29	技術員 把 [零件/傳票] 有秩序的 [檢查過/寄出去] 了以後，就放心了。	The technician BA orderly the parts/summons inspected/mailed and then felt assured.
30	商人 把 [訂單/聽診器] 毫無保留的 [檢查好/燒掉] 了以後，就去打球了。	The businessman BA thoroughly the order sheet/stethoscope checked/burned and then went to play golf.
31	收銀員 把 [發票/粉筆] 一轉眼 [丟掉/晾乾] 了以後，就挨罵了。	The cashier BA quickly the receipt/chalk threw away/dried and was scolded.
32	鋼琴家 把 [樂譜/炸彈] 終於 [放好/挖出來] 了以後，就去渡假了。	The pianist BA finally the music score/bomb properly-placed/dug up and then went on vacation.
33	醫生 把 [聽診器/麥克風] 嚴密的 [消毒好/鎖起來] 了以後，就去喝茶了。	The doctor BA thoroughly the stethoscope/microphone disinfected/locked away and then went to have tea.
34	運動員 把 [獎杯/樂譜] 生氣的 [裝箱/丟掉] 了以後，就回家了。	The athlete BA angrily the trophy/music score packed/got rid of and then went home.
35	牛仔 把 [靴子/設計圖] 著急的 [烘乾/剪掉] 了以後，就去騎馬了。	The cowboy BA hurriedly the boots/design paper dried/cut up and then went horse riding.
36	書法家 把 [毛筆/草藥] 輕鬆的 [晾乾/消毒好] 了以後，就去吃喝茶了。	The calligrapher BA gently the paintbrush/herbs air-dried/disinfected and then went to have tea.
37	作曲家 把 [鋼琴/零件] 心甘情願的 [調整好/拆掉] 了以後，就去散步了。	The songwriter BA willingly the piano/mechanic parts adjusted/dismantled and then went for a walk.
38	伴娘 把 [婚紗/獎杯] 痛心的 [收起來/丟掉] 了以後，就去睡了。	The bridesmaid BA sadly the wedding dress/trophy put away/threw away and then went to sleep.
39	老師 把 [粉筆/擔架] 霸道的 [弄斷/賣掉] 了以後，就去上課了	The teacher BA ruthlessly the chalk/stretcher broke apart/sold and then went to class.
40	獵人 把 [槍枝/氧氣瓶] 慌忙的 [測試過/沒收] 了以後，就出發了。	The hunter BA hurriedly the gun/oxygen bottle tested/confiscated and then set out.
41	調酒師 把 [酒杯/背包] 默默的 [淘汰/審核完] 了以後，就去工作了。	The bartender BA quietly the wine glass/backpack got rid of/approved and then went to work.
42	導遊 把 [特產/竹籃] 令人吃驚的 [推銷出去/藏起來] 了以後，就發財了。	The tour guide BA surprisingly the local specialties/bamboo basket sold/hid and then became rich.
43	漁夫 把 [魚網/勳章] 奮力的 [扔出去/割壞] 了以	The fisherman BA forcefully the

	後，就上船了。	fishnet/medal threw away/cut and then got on the boat.
44	作家 把 [小說/簽證] 緊張的 [核對完/燒掉] 了之後，就出國了。	The writer BA nervously the novel/visa audited/burned and then went abroad.
45	看護 把 [輪椅/鋼琴] 嚴格的 [擦乾淨/買下來] 了以後，就去上班了。	The caretaker BA carefully the wheelchair/piano wiped cleaned/bought and then went to work.
46	建築工人 把 [水泥/婚紗] 有技巧的 [晾乾/染色] 了以後，就去喝酒了。	The construction worker BA skillfully the cement/wedding gown dried/dyed and then went for a drink.
47	投資人 把 [股票/鏡頭] 自顧的 [賣掉/交出去] 了以後，就去渡假了。	The investor BA secretly the stock options/camera lens sold/handed in and then went on vacation.
48	移民官 把 [簽證/撲克牌] 憤怒的 [沒收/撕掉] 了以後，就離開了。	The immigration officer BA angrily the visa/playing cards confiscated/tore up and then left.

APPENDIX F: EXPERIMENTAL STIMULI FOR ONLINE SVO SELF-PACED READING STUDY

The stimuli for the online SVO self-paced ready study are presented below. There were 48 sets of stimuli with four conditions in each set. The verb was manipulated and was either a typical or atypical action performed by the protagonist. The typical verb and atypical verb are separated by a slash line (/), with the typical verb presented first. The object noun was manipulated and was either strongly or weakly associated with the subject noun. The two object nouns in each set are presented after the verbs and separated by a slash line (/), with the strongly associated noun being presented first.

For example, in set 1, *chef* is the subject noun. The typical and atypical verbs for this set were *washed* and *mailed,* respectively. The two object nouns used were *knife* and *military uniform* which were strongly and weakly associated respectively. All the stimuli items were presented in Chinese and presented as SVO sentences and embedded into a longer sentence to allow for more natural reading.

1	廚師 [清洗好/寄出去] 了 [刀子/迷彩服] 以後，就下班了。	The chef washed/mailed the knife/military uniform and then got off work.
2	演員 [打開/檢查好] 了 [劇本/烤箱] 以後，就離開了。	The actor opened/checked the script/oven and then left.
3	郵差 [送出/收下] 了 [包裹/警棍] 以後，就離開了。	The postman delivered/collected the package/baton and then left.
4	警察 [清理好/遺失] 了 [警車/培養皿] 以後，就出去了。	The policeman cleaned/lost the police car/petri dish and then went outside.
5	服務生 [收下/退回] 了 [小費/紀錄片] 以後，就跑掉了。	The waiter accepted/returned the tip/documentary and then ran away.
6	公車司機 [檢查過/偷走] 了 [車票/刀子] 以後，就心安了。	The bus driver checked/stole the ticket/the knife and then felt assured.
7	秘書 [檢查好/烘乾] 了 [文件/顯微鏡] 以後，就放心了。	The secretary checked/blow-dried the document/microscope and then felt assured.
8	捕手 [弄破/消毒好] 了 [手套/放大鏡] 之後，就去休息了。	The catcher broke/disinfected the glove/magnifying glass and then went to take a rest.
9	農夫 [做好/寄出] 了 [稻草人/簡報] 以後，就收工	The farmer prepared/sent the

了。

scarecrow/presentation slides and then packed up.

10 富翁 [買下/清洗好] 了 [別墅/口哨] 之後，就很後悔。

The millionaire bought/cleaned the mansion/whistle and then regretted it.

11 畫家 [買下/鎖起] 了 [畫筆/程式] 之後，就開工了。

The painter bought/locked away the paintbrush/software program and then started to work.

12 鎖匠 [破壞/封起] 了 [門鎖/雜誌] 之後，就跑走了。

The locksmith destroyed/sealed the door lock/magazine and then ran away.

13 難民 [收下/保養好] 了 [毛毯/紀念品] 之後，就走掉了。

The refugee accepted/maintained the blanket/souvenir and then walked away.

14 學生 [弄丟/賣掉] 了 [書包/古董] 以後，就很憂心。

The student lost/sold the school bag/antique and then felt worried.

15 毒販 [製作好/沒收] 了 [毒品/履歷表] 以後，就躲起來了。

The drug dealer made/confiscated the drug/resume and then went to hide.

16 獸醫 [準備好/推銷出] 了 [麻醉藥/救生艇] 以後，就去休息了。

The vet prepared/sold the anesthesia/lifeboat and then went to rest.

17 消防員 [檢查好/鎖起] 了 [滅火器/展覽品] 以後，就放心了。

The fireman inspected/locked away the fire extinguisher/exhibition item and then felt assured.

18 教授 [寄出/晾乾] 了 [推薦信/滾筒] 以後，就去上課了。

The professor mailed/dried the recommendation letter/roller and then went to class.

19 保鏢 [套上/做好] 了 [西裝/假牙] 以後，就去上班了。

The bodyguard put on/made the suit/denture and then went to work.

20 探險家 [遺失/燒掉] 了 [指南針/高跟鞋] 以後，就很困擾。

The explorer lost/burned the compass/heels and felt worried.

21 考古學家 [挖出/破壞] 了 [古蹟/毒品] 以後，就離開了。

The archeologist excavated/destroyed the historical site/drug and then left.

22 裁判 [擦乾淨/拍照] 了 [口哨/包裹] 以後，就去工作了。

The referee cleaned/photographed the whistle/package and then went to work.

23 小嬰兒 [打翻/打開] 了 [奶瓶/葡萄酒] 以後，就爬走了。

The baby spilled/opened the milk bottle/wine and then crawled away.

24 觀光客 [挑選好/丟掉] 了 [紀念品/油漆] 以後，就去拍照了。

The tourist selected/threw away the souvenir/paint and then went to take a picture.

25 救生員 [換掉/拍賣] 了 [急救箱/手套] 以後，就去巡察了。

The lifeguard changed/auctioned the first aid kit/glove and then went to patrol.

26 生物學家 [消毒過/破壞] 了 [培養皿/別墅] 以後，就休假去了。

The biologist disinfected/destroyed the petri dish/mansion and then went on vacation.

27 小孩 [吞下/製作好] 了 [棒棒糖/發票] 以後，就跑走了。

The child swallowed/made the lollipop/receipt and then ran away.

28 街頭藝人 [撿起/丟出] 了 [銅板/救生衣] 以後，就收工了。

The street performer picked up/tossed out the coins/life vest and then packed up.

29	技術員 [檢查過/寄出] 了 [零件/傳票] 以後，就放心了。	The technician inspected/mailed the parts/summons and then felt assured.
30	商人 [檢查好/燒掉] 了 [訂單/聽診器] 以後，就去打球了。	The businessman checked/burned the order sheet/stethoscope and then went to play golf.
31	收銀員 [丟掉/晾乾] 了 [發票/粉筆] 以後，就挨罵了。	The cashier threw away/dried the receipt/chalk and was scolded.
32	鋼琴家 [放好/挖出] 了 [樂譜/炸彈] 以後，就去渡假了。	The pianist properly-placed/dug up the music score/bomb and then went on vacation.
33	醫生 [消毒好/鎖起] 了 [聽診器/麥克風] 以後，就去喝茶了。	The doctor disinfected/locked away the stethoscope/microphone and then went to have tea.
34	運動員 [裝箱/丟掉] 了 [獎杯/樂譜] 以後，就回家了。	The athlete packed/got rid of the trophy/music score and then went home.
35	牛仔 [烘乾/剪掉] 了 [靴子/設計圖] 以後，就去騎馬了。	The cowboy dried/cut up the boots/design paper and then went horse riding.
36	書法家 [晾乾/消毒好] 了 [毛筆/草藥] 以後，就去吃喝茶了。	The calligrapher air-dried/disinfected the paintbrush/herbs and then went to have tea.
37	作曲家 [調整好/拆掉] 了 [鋼琴/零件] 以後，就去散步了。	The songwriter adjusted/dismantled the piano/mechanic parts and then went for a walk.
38	伴娘 [收起/丟掉] 了 [婚紗/獎杯] 以後，就去睡了。	The bridesmaid put away/threw away the wedding dress/trophy and then went to sleep.
39	老師 [弄斷/賣掉] 了 [粉筆/擔架] 以後，就去上課了。	The teacher broke apart/sold the chalk/stretcher and then went to class.
40	獵人 [測試過/沒收] 了 [槍枝/氧氣瓶] 以後，就出發。了。	The hunter tested/confiscated the gun/oxygen bottle and then set out.
41	調酒師 [淘汰/審核完] 了 [酒杯/背包] 以後，就去工作了。	The bartender got rid of/approved the wine glass/backpack and then went to work.
42	導遊 [推銷出/藏起] 了 [特產/竹籃] 以後，就發財了。	The tour guide sold/hid the local specialties/bamboo basket and then became rich.
43	漁夫 [扔出/割壞] 了 [魚網/勳章] 以後，就上船了。	The fisherman threw away/cut the fishnet/medal and then got on the boat.
44	作家 [核對完/燒掉] 了 [小說/簽證] 之後，就出國了。	The writer audited/burned the novel/visa and then went abroad.
45	看護 [擦乾淨/買下] 了 [輪椅/鋼琴] 以後，就去上班了。	The caretaker wipe cleaned/bought the wheelchair/piano and then went to work.
46	建築工人 [晾乾/染色] 了 [水泥/婚紗] 以後，就去喝酒了。	The construction worker dried/dyed the cement/wedding gown and then went for a drink.
47	投資人 [賣掉/交出] 了 [股票/鏡頭] 以後，就去渡假了。	The investor sold/handed in the stock options/camera lens and then went on

		vacation.
48	移民官 [沒收/撕掉] 了 [簽證/撲克牌] 以後，就離開了。	The immigration officer confiscated/tore up the visa/playing cards and then left.

References

Altmann, G., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*(4), 502-518.

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73,* 247–264.

Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin & Review*, *4*(4), 439-461.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language,* 59(4), 390-412.

Bicknell, K., Elman, J. L., Hare, M., McRae, K., & Kutas, M. (2010). Effects of event knowledge in processing verbal arguments. *Journal of Memory and Language, 63,* 489–505. [doi:10.1016/j.jml.2010.08.004]

Bird, H., Howard, D., & Franklin, S. (2003). Verbs and nouns: The importance of being imageable. *Journal of Neurolinguistics*, *16*(2), 113-149.

Bornkessel, I., & Schlesewsky, M. (2006). The extended argument dependency model: A neurocognitive approach to sentence comprehension across languages. *Psychological Review, 113,* 787–821.

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. PLoS One, 5(6), e10729.

Camblin, C. C., Gordon, P. C., & Swaab, T. Y. (2007). The interplay of discourse congruence and lexical association during sentence processing: Evidence from ERPs and eye-tracking. Journal of Memory and Language, 56, 103–128.

Chao, Y. R. (1968). *A grammar of spoken Chinese.* Berkeley, CA: University of California Press.

Chomsky, N. (1965). *Aspects of the Theory of Syntax* (No. 11). MIT press.

Chu, C. C., & Chi, T-J. (2006). *A Cognitive-Functional Grammar of Mandarin Chinese.* Taipei, The Crane Publishing Co.

Cook, A. E., & Myers, J. L. (2004). Processing discourse roles in scripted narratives: The influences of context and world knowledge. *Journal of Memory and Language*, *50*(3), 268-288.

Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and cognitive processes*, *13*(1), 21-58.

Dowty, D. (1991). Thematic proto-rules and argument selection. *Language, 67*, 547-619.

Drews, E., & Zwitserlood, P. (1995). Morphological and orthographic similarity in visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(5), 1098.

Drummond, A. (2010). IBEX (Version 0.3.6) [Software]. Available from http://spellout.net/ibexfarm/

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language, 41,* 469–495.

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain research*,*1146*, 75-84.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive psychology,* 47(2), 164-203.

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory & Language, 25,* 348-368.

Ferretti, T. R., McRae, K., & Hatherell, A. (2001). Integrating verbs, situation schemas, and thematic role concepts. Journal of Memory and Language, 44(4), 516-547.

Filik, R. (2008). Contextual override of pragmatic anomalies: Evidence from eye movements. *Cognition*, *106*(2), 1038-1046.

Filik, R., & Leuthold, H. (2008). Processing local pragmatic anomalies in fictional contexts: Evidence from the N400. *Psychophysiology*, *45*(4), 554-558.

Fischler, I. (1977). Semantic facilitation without association in a lexical decision task. *Memory & Cognition*, *5*(3), 335-339.

Fischler, I., Bloom, P. A., Childers, D. G., Roucos, S. E., & Perry, N. W. (1983). Brain potentials related to stages of sentence verification. *Psychophysiology*, *20*(4), 400-409.

Frazier, L. (1995). Constraint satisfaction as a theory of sentence processing.*Journal of Psycholinguistic Research*, *24*(6), 437-468.

Goldin-Meadow, S., Seligman, M. E., & Gelman, R. (1976). Language in the two-year old. *Cognition, 4*(2), 189-202.

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science, 304*. 438-441.

Hare, M., Jones, M., Thomson, C., Kelly, S., & McRae, K. (2009). Activating event knowledge. *Cognition 111,* 151–167.

Hoeks, J. C., Stowe, L. A., & Doedens, G. (2004). Seeing words in context: the interaction of lexical and sentence level information during reading. *Cognitive Brain Research*, *19*(1), 59-73.

Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychonomic Bulletin & Review*, *10*(4), 785-813.

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford University Press.

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General, 111,* 228-238.

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language, 49,* 133–156.

Kamide, Y., Scheepers, C., & Altmann, G. T. M. (2003). Integration of syntactic and semantic information in predictive processing: Cross- linguistic evidence from German and English. *Journal of Psycholinguistic Research, 32,* 37–55.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and cognitive processes*, *15*(2), 159-201.

Keller, F., Gunasekharan, S., Mayo, N., & Corley, M. (2009). Timing accuracy of web experiments: A case study using the WebExp software package. *Behavior Research Methods*, *41*(1), 1-12.

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: the role of working memory. *Journal of Memory and Language, 30.* 580-602.

Kuperberg, G. R., Kreher, D. A., Sitnikova, T., Caplan, D. N., & Holcomb, P. J. (2007). The role of animacy and thematic relationships in processing active English sentences: Evidence from event-related potentials. *Brain and language*,*100*(3), 223-237.

Kuperberg, G. R., Holcomb, P. J., Sitnikova, T., Greve, D., Dale, A. M., & Caplan, D. (2003). Distinct patterns of neural modulation during the processing of conceptual and syntactic anomalies. *Journal of Cognitive Neuroscience*,*15*(2), 272-293.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203-205.

Li, C. N., & Thompson, S. A. (1981). *Mandarin Chinese: A Functional Reference Grammar*. Berkeley: University of California Press.

Li, P., Bates, E., & MacWhinney, B. (1993). Processing a Language without Inflections: A Reaction Time Study of Sentence Interpretation in Chinese. *Journal of Memory and Language, 32,* 2, Apr, 169-192.

Lupker, S. J. (1984). Semantic priming without association: A second look. *Journal of Verbal Learning and Verbal Behavior*, *23*(6), 709-733.

Ma, W., Golinkoff, R. M., Hirsh-Pasek, K., McDonough, C., & Tardif, T. (2009). Imageability predicts the age of acquisition of verbs in Chinese children. *Journal of Child Language*, *36*(02), 405-423.

Markman, E. M., & Hutchinson, J. E. (1984). Children's sensitivity to constraints on word meaning: Taxonomic versus thematic relations. *Cognitive psychology*, *16*(1), 1-27.

Marslen-Wilson, W. D. (1975). Sentence perception as an interactive parallel process. *Science, 189,* 226-228.

Matsuki, K., Chow, T., Hare, M., Elman, J. L., Scheepers, C., & McRae, K. (2011). Event-based plausibility immediately influences on-line language comprehension. *Journal of Experimental Psychology 37,* No. 4, 913–934. [DOI: 10.1037/a0022964]

McRae, K. & Matsuki, K. (2009). People use their knowledge of common events to understand language, and do so as quickly as possible. *Language & Linguistics Compass, 3,* 1417-1429.

McRae, K., Hare, M., Elman, J. L., & Ferretti, T. R. (2005). A basis for generating expectancies for verbs from nouns. *Memory and Cognition,* 33, 1174–1184.

McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, *38*(3), 283-312.

Metusalem, R., Kutas, M., Urbach, T. P., Hare, M., McRae, K., & Elman, J. L. (2012). Generalized event knowledge activation during online sentence comprehension. *Journal of memory and language,* 66(4), 545-567.

Mitchell, D. C. (1984). An evaluation of subject-paced reading tasks and other methods for investigating immediate processes in reading. *New methods in reading comprehension research*, 69-89.

Mitchell, D. C. (2004). On-line methods in language processing: Introduction and historical review. In M. Carreiras, & C. Clifton, Jr. (Eds.), *The On-Line Study of Sentence Comprehension: Eye-tracking, ERPs and Beyond* (pp. 15-32). New York: Psychology Press.

Murray, W. S. (2006). The nature and time course of pragmatic plausibility effects. *Journal of psycholinguistic research*, *35*(1), 79-99.

Nieuwland, M. S., & Van Berkum, J. J. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of cognitive neuroscience, 18*(7), 1098-1111.

Osterhout, L., & Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and language*, *34*(6), 739-773.

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of memory and language*, *31*(6), 785-806.

Pecher, D., Zeelenberg, R., & Raaijmakers, J. G. (1998). Does pizza prime coin? Perceptual priming in lexical decision and pronunciation. *Journal of Memory and Language*, *38*(4), 401-418.

Philipp, M., Bornkessel-Schlesewsky, I., Bisang, W., & Schlesewsky, M**.** (2008). The role of animacy in the real time comprehension of Mandarin Chinese: Evidence from auditory event-related brain potentials. *Brain and Language* 105, 112-133.

Rayner, K., Warren, T., Juhasz, B. J., & Liversedge, S. P. (2004). The effect of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(6), 1290.

Rayner, K., Carlson, M., & Frazier, L. (1983). The inteaction of syntax and semantics during sentene processing: eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learnng and Verbal Behavior, 22.* 358-374.

Schutze, C. T. (1996). *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. University of Chicago Press.

Spence, D. P., & Owens, K. C. (1990). Lexical co-occurrence and association strength. *Journal of Psycholinguistic Research*, *19*(5), 317-330.

Staub, A., & Rayner, K. (2007). Eye movements and on-line comprehension processes. *The Oxford handbook of psycholinguistics*, *327*, 342.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268,* 1632-1634.

Tanenhaus, Michael K., Greg Carlson, and John C. Trueswell. "The role of thematic structures in interpretation and parsing." *Language and cognitive processes* 4.3-4 (1989): SI211-SI234.

Tardif, T. (1996). Nouns are not always learned before verbs: Evidence from Mandarin speakers' early vocabularies. *Developmental Psychology*, *32*(3), 492.

Traxler, M. J., Morris, R. K. & Seely, R. E. (2002). Processing subect and object relative clauses: evidence from eye movements. *Jornal of Memory and Langauge, 47.* 69-90.

Tyler, L. K., Bright, P., Fletcher, P., & Stamatakis, E. A. (2004). Neural processing of nouns and verbs: The role of inflectional morphology.*Neuropsychologia*, *42*(4), 512-523.

Van Gompel, R. P., Pickering, M. J., Pearson, J., & Liversedge, S. P. (2005). Evidence against competition during syntactic ambiguity resolution. *Journal of Memory and Language*, *52*(2), 284-307.

Van Berkum, J. J., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(3), 443.

Warren, T., McConnell, K., & Rayner, K. (2008). Effects of context on eye movements when reading about possible and impossible events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(4), 1001.

Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications. arXiv:1308.5499.

Whitlock, M. C., & Schluter, D. (2009). *The analysis of biological data*. Greenwood Village, Colorado: Roberts and Company Publishers.

Zarcone, A., & Padó, S. (2011). Generalized event knowledge in logical metonymy resolution. In *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 944-949).